12th INTERNATIONAL CONGRESS ON ACOUSTICS

12e CONGRÈS INTERNATIONAL D'ACOUSTIQUE

12. INTERNATIONALER KONGRES FÜR AKUSTIK

VOLUME / BAND I
A - C

TORONTO, CANADA
24-31 JULY 1986
12th INTERNATIONAL CONGRESS ON ACOUSTICS

12e CONGRÈS INTERNATIONAL D'ACOUSTIQUE

12. INTERNATIONALER KONGRES FÜR AKUSTIK

VOLUME/BAND I
A - C

TORONTO, CANADA
24-31 JULY 1986
HOW TO USE THESE PROCEEDINGS

Each paper is designated by a letter, A to M, and two numbers. The same designation is used also for the Abstract of the paper where it appears in the Program.

The letter designates the Subject Classification as listed below. The first of the two numbers designates the sequence of the Technical Sessions in each subject classification, and represents the Session to which the paper was allocated. The second of the two numbers designates the ordering of the several papers in the Technical Session.

These Proceedings are bound in three volumes. The papers are printed in alphanumeric sequence starting with paper A1-1 near the beginning of Volume 1 and finishing with paper M4-8 near the end of Volume 3. The complete Author index is repeated at the end of each volume of the Proceedings and also appears at the end of the Program.

COMMENT UTILISER CES VOLUMES

Chaque communication est classée selon une lettre, de A à M, et deux numéros. Ce même classement accompagne le résumé qui se trouve dans le programme.

La lettre reprend la Classification des thèmes ci-dessous. Le premier numéro repère l'ordre des séances dans chaque thème et indique la séance dans laquelle se trouve la conférence. Le deuxième numéro repère l'ordre des conférences dans chaque séance.

Ces actes sont reliés en trois volumes. Les communications sont imprimées par ordre alphabétique, commençant par la conférence A1-1 au début du 1er volume et se terminant par la conférence M4-8 à la fin du 3e volume. Une liste complète de tous les auteurs se trouve à la fin de chaque volume et aussi à la fin du programme.

ANLEITUNG ZUR BENUTZUNG DER KONGRESSBERICHTE

Jeder Beitrag wird durch einen Buchstaben, A bis M, und durch zwei Zahlen gekennzeichnet. Die gleiche Bezeichnung wird auch für die Kurzfassung des Beitrags im Tagungsprogramm benutzt.

Der Buchstabe bezieht sich auf die weiter unten angegebene Themenklassifikation der Fachgebiete. Die erste der zwei Zahlen bezieht sich auf die Anordnung der Technischen Sitzungen in dem gewissen Fachgebiet und kennzeichnet diejenige Sitzung, welcher der bestimmte Beitrag zugeteilt wurde. Die zweite der beiden Zahlen bezieht sich auf die Reihenfolge der verschiedenen Beiträge in der bestimmten Technischen Sitzung.


<table>
<thead>
<tr>
<th>SUBJECT CLASSIFICATION</th>
<th>CLASSIFICATION DES CATEGORIES</th>
<th>THEMENKLASSEIFIKATION</th>
<th>VOLUME/BAND</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Speech Communication</td>
<td>Communication parlée</td>
<td>Sprachkommunikation</td>
<td>I</td>
</tr>
<tr>
<td>B Physiological and Psychological Acoustics</td>
<td>Physio-Psycho-acoustique</td>
<td>Physiologische und Psychologische akustik</td>
<td>I</td>
</tr>
<tr>
<td>C Noise</td>
<td>Bruit</td>
<td>Lärm</td>
<td>I</td>
</tr>
<tr>
<td>D Shock and Vibration</td>
<td>Choc et vibrations</td>
<td>Schwingungen und Erschütterung</td>
<td>II</td>
</tr>
<tr>
<td>E Architectural and Building Acoustics</td>
<td>Acoustique architecturale et du bâtiment</td>
<td>Architektur und Gebäudeakustik</td>
<td>II</td>
</tr>
<tr>
<td>F Bioacoustics</td>
<td>Bio-acoustique</td>
<td>Biologische Akustik</td>
<td>II</td>
</tr>
<tr>
<td>G Ultrasonics, Quantum and Physical Effects</td>
<td>Ultrasons, Quanta, Effets physiques</td>
<td>Ultraschall, Quanten und Physikalische Einflüsse</td>
<td>II</td>
</tr>
<tr>
<td>H Underwater Acoustics</td>
<td>Acoustiques sous-marine</td>
<td>Wasserschall</td>
<td>III</td>
</tr>
<tr>
<td>I Physical Acoustics</td>
<td>Acoustique physique</td>
<td>Physikalische Akustik</td>
<td>III</td>
</tr>
<tr>
<td>J Aerocoustics and Atmospheric Sound</td>
<td>Aéroacoustique, Acoustique atmosphérique</td>
<td>Aero-Akustik und Schall in der Atmosphäre</td>
<td>III</td>
</tr>
<tr>
<td>L Transduction</td>
<td>Transduction</td>
<td>Schallwandler</td>
<td>III</td>
</tr>
<tr>
<td>M Measurements, Instrumentation, Signal Processing, and Statistical Methods</td>
<td>Mesures et instruments, Traitement du signal, Méthodes statistiques</td>
<td>Messung, Instrumente, Signal Auswertung, Statistische Verfahren, Plenarvorträge</td>
<td>III</td>
</tr>
<tr>
<td>Plenary Lectures</td>
<td>Conférences plénières</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
These Proceedings are published on behalf of the Technical Program Committee by the Executive Committee of the 12th International Congress on Acoustics, a Committee of the Canadian Acoustical Association. The Technical Program Committee served as the Editorial Board for the Congress Proceedings and Program.

Technical Program Committee:
T.F.W. Embleton, Chairman
G.A. Daigle
M.R. Stinson
A.C.C. Warnock
assisted by
J. Quaroni
M.M. Vaillancourt

Technical Subject Coordinators:
A — M. Hunt
B — R. Hétu, N. Lalande
C — G. Krishnappa
D — H. Davies
E — J.S. Bradley, J.D. Quirt
F — A.J. Brammer, R. Clarke
G — D. Hutchins
H — G.H. Brooke
I — L.T. Russell
J — W.T. Chu
K — A. Cohen
L — G.S.K. Wong
M — J. Nicolas

Permission is hereby granted for any person to reproduce a fractional part of any paper herein, provided that permission is obtained from its author(s) and credit is given to the author(s) and these Proceedings. An author, or his research sponsor, may reproduce his paper in full, crediting these Proceedings; this permission is not assignable.
SUBJECTIVE SNR MEASURE FOR QUALITY ASSESSMENT OF SPEECH CODERS - A CROSS LANGUAGE STUDY

Mamoru NAKATSUI

Radio Research Laboratory, Ministry of Posts and Telecommunications, 4-2-3, Nukuiita, Koganei-shi, Tokyo, Japan 184

INTRODUCTION

The subjective assessment method expected to provide a practical engineering criterion for overall speech quality should satisfy the following requirements of engineering practice:

1. Test administration and data reduction are simple enough to be carried out in most speech-communication laboratories, without assistance of specialist in psychological experiments or of trained or professional listeners.

2. Reproducibility of the results across studies is high enough to enable one to compare the measures obtained on different occasions or at different laboratories.

The subjective speech-to-noise ratio (SNR) has been proposed as one of such measures and experimental results of subjective tests with the measure have been reported [1]. The test results show that

1. The subjective SNR measure provides an adequate singe absolute standard for a wide range of speech waveform coders,

2. Reliable score is available using reasonably small number of listeners and speakers, and

3. No significant speaker and listener variation is found in the scores of two separate test sessions 14 months apart using different groups of English speakers and listeners.

The purpose of this study is to investigate whether reproducibility of the results across tests is high enough to enable one to compare directly the measures obtained at different laboratories having different nationalities and language backgrounds.

SUBJECTIVE SNR (EQUIVALENT Q) MEASURE

The concept of a subjective SNR has been found in the iso-preference method originally introduced by Mansou and Kasliu [2]. The subjective SNR is derived from the forced-choice pair-comparison test using the psychometric analysis procedure commonly used in the method of constants. A speech signal degraded by varying amounts of multiplicative white noise [3] is selected as the reference system in our tests.

Reference System

The sampled reference signal \( r(t) \) corrupted by multiplicative white noise \( n(t) \) is defined as

\[
\begin{align*}
r(t) &= \{a(t) + n(t)\}/\sqrt{a_n}, \quad (la) \\
n(t) &= a \cdot n(t) - e(t), \quad (lb)
\end{align*}
\]

where \( a(t) \) is original speech signal which is also served as input signal to the speech coders evaluated, \( a \) is a coefficient for SNR control, and \( e(t) \) is a random variable taking on values of \( +1 \) and \( -1 \) with equal probability and independently of \( a(t) \) for each sample. Since this reference signal is identical to one of the speech signals processed by the modulated-noise reference unit (MNHU) of CCITT and the SNR of the speech signal degraded by MNHU is called Q, our subjective SNR measure can be called as Equivalent Q measure.

0.5 sec. A B A B 4 sec. A B

Fig. 1. Time pattern of the test pair sequence.

\[
\text{RATE, PREFERENCES TEST SIGNAL}
\]

\[
\text{SNR OF REFERENCE SIGNAL IN dB}
\]

\[
25 \quad 19 \quad 16 \quad 13
\]

\[
0.0 \quad 0.2 \quad 0.4 \quad 0.6 \quad 0.8 \quad 1.0
\]

Fig. 2. An example of the least-square fit of the preference data: black circles denote the rates preferring the test signal and a solid curve denote the normally fitted ogive.

Subjective Test Format

Each test signal to be evaluated is paired with five or six reference signals selected so that preference ranging from 0% to 100% would result. During the test, the listeners are presented with repeated signal pairs in the order of ABAB as shown in Fig. 1. The listeners are asked to mark as a source of information.

Test Data Analysis

The proportion of listeners preferring the test signal \( p(t) \) is converted to unit normal deviate \( z(t) \) using the equation

\[
p(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z(t)} e^{-(1/2)t^2} dt. \quad (2)
\]

Applying Muller-Urban weighting to the converted data, a weighted least square algorithm [4] is used to fit a straight line to the data points:

\[
z = a + b \cdot s,
\]

where \( s \) represents SNR of the reference signal. Thus the subjective SNR (SNR_subj) and its standard deviation are given by

\[
\text{SNR_subj} = -a/b, \quad (4a)
\]

\[
s = 1/b. \quad (4b)
\]

Fig. 2 shows an example of the least-square fit of the preference data. The 95% confidence interval of SNR_subj as an estimate of population mean \( \mu \) is defined by

\[
|\text{SNR_subj} - \mu| < t(\phi, z) \sigma/\sqrt{\phi}, \quad (5)
\]

where \( t(\phi, z) \) is the \( \phi \) distribution with \( z \) degrees of freedom and \( x \) is the significance level (5%).

EXPERIMENTAL PROCEDURE

Table 1 summarizes the experimental setups of the current test (Test III) and the previous tests (Tests I and II). Five kinds of coder configura
Table 1. Experimental Frameworks

<table>
<thead>
<tr>
<th>Experimental Factors</th>
<th>Test I</th>
<th>Test II</th>
<th>Test III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Place of the Test</td>
<td>Montreal</td>
<td>Montreal</td>
<td>Tokyo</td>
</tr>
<tr>
<td>Date of the Test</td>
<td>Mar '79</td>
<td>May '80</td>
<td>Nov '85</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
<td>English</td>
<td>Japanese</td>
</tr>
<tr>
<td>Number of Utterances</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Number of Speakers</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Number of Sentences</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Number of Listeners</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Number of Coders</td>
<td>9</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Total Number of Pairs</td>
<td>392</td>
<td>210</td>
<td>240</td>
</tr>
</tbody>
</table>

-tons were simulated and evaluated in the current test, including 40 and 56 kb/s log-PCMs, 16 and 32 kb/s dual-adaptive delta modulators (DAMs) [5], and a 16 kb/s adaptive delta modulator with one-bit memory (ADM) [6]. Two PCM and a ADM configurations were used as anchor points connecting the current test in Japan with the previous tests in Canada.

Four short sentences spoken by two male and two female speakers were bandlimited from 200 to 3400 Hz and digitized to 12 bits at sampling frequencies of 8, 16, and 32 kHz. Those digitized speech samples were served as inputs to the two PCM, 32 kb/s DAM, and 16 kb/s ADM configurations to produce the test signals to be evaluated. The speech samples having non-standard bandlimitation from 200 to 2400 Hz were served as inputs to the 16 kb/s DAM and the 40 kb/s PCM again. The reference signals were also processed by Eq. (1) using these speech samples.

Two separate sessions of pair comparison tests, one session using speech samples of standard bandwidth and the other session using those of non-standard (narrow) bandwidth, were conducted with eleven untrained listeners. Participants in both sessions were native Japanese speakers.

RESULT AND DISCUSSION

A subjective SNR of each test signal and its 95% confidence interval were estimated from the test data pooled over four utterances and all the listeners and shown in Fig. 3 together with the results of the previous tests. Estimates obtained from the speech data having non-standard (narrow) bandwidth are indicated by an arrow in the figure. No statistically significant differences between the subjective SNR estimates at the 5% level is found for the following test signal pairs: (1) 40 kb/s PCM in tests I, II, and III, (2) the 36 kb/s PCM in tests I and III, and (3) 16 kb/s ADM in tests I and III. The subjective SNR measure gives quite reliable scores evaluated in different laboratories having different language background. No significant variation due to the factors of speaker and listener is found for the test signals of the current test as has also been shown in the previous tests for the waveform coders.

The subjective SNR measure, which is the absolute scale derived from relative judgements, truly satisfies the requirements of engineering practice described before. On the other hand, a mean opinion score (MOS), which is the most widely used subjective measure on overall speech quality and is the absolute scale derived from absolute (categorical) judgements, has shown remarkable variations in test results obtained in different countries for the same speech transmission system [7]. A MOS Equivalent Q measure [8] aiming at interpreting a large set of MOS data-base pooled into the subjective SNR scale, has shown saturation (non-linearity) in high quality range (56 to 64 kb/s PCM) reflecting inadequacy of the MOS measure.

REFERENCES

INTRODUCTION

L'influence de la langue sur le spectre moyen à long terme (SMLT) est mal connue. D'un point de vue théorique, il pourrait être considéré que le SMLT, réputé indépendant du contenu du signal de parole, ne varie pas lorsqu'un locuteur s'exprime dans plusieurs langues. Nous avons cependant montré [1] que le SMLT n'est pas que relativement indépendant du contenu phonémique. A fortiori, il semblerait donc logique qu'il varie aussi sous l'effet de variations des langues parlées.

La littérature ne permet pas de statuer définitivement sur cette hypothèse. Malemski et Hollien [2], d'une part, ont appliqué aux productions vocales de deux échantillons indépendants de sujets polonais et américains les mêmes techniques d'identification à base de SMLT. Ayant obtenu des scores différents à partir des deux langues, ils supposent l'existence d'effets différenciels liés aux langues elles-mêmes. Tosi [3], d'autre part, a étudié les spectres choréus provenant des productions françaises, italiennes et pictoises de sujets trilingues. Il conclut, quant à lui, à une existence de "invariances relatives" dans leurs spectres. Comme le remarque Nolan [4], il existe un conflit (ou soulis) indirect entre ces conclusions plaçant tantôt en faveur d'effets "langues" prêgnants, tantôt en faveur d'effets "sujets" prêgnants.

Dans la présente recherche, nous tentons d'étudier ces effets à partir des SMLT de sujets bilinéaires exprimant en français et en néerlandais.

EXPERIMENTATION

Dispositif expérimental

Les locuteurs étaient 10 hommes francophones belges âgés de 18 à 21 ans et habitués à s'exprimer spontanément en néerlandais. Chacun fut invité à répéter 10 fois un texte français ainsi qu'un texte néerlandais phonétiquement équilibré. Ces textes étaient d'une durée approximative de 18 secondes.

Les productions furent enregistrées au cours d'une seule séance pour chaque sujet, sur magnétophone Magra IV S, via un microphone Neuman RH 84. Les enregistrements se déroulèrent dans un studio insonorisé, le micro étant maintenu à distance constante (40 cm) des lèvres du sujet. Les enregistrements français ainsi obtenus ont fait l'objet d'études antérieures [1].

Traitement acoustique

Les analyses acoustiques furent réalisées ultérieurement au moyen de l'analyseur spectral Bruel & Kjaer 2033. Celui-ci, échantillonnant le signal d'entrée à 12,8 KHz, produit des spectres définis sur 400 canaux dans la bande 0-5 KHz, avec une résolution constante de 12,5 Hz. Ces spectres instantanés furent appliqués à l'entrée d'un algorithme de calcul de moyennes à pondération linéaire qui fournit, pour chaque production, un SMLT. Les 200 (10 sujets x 2 langues x 10 productions) SMLT ainsi obtenus furent ensuite transmis à un ordinateur Apple II via un interface GPIB en vue du traitement statistique.

Traitement statistique

Afin de rendre les résultats de la présente expérience compatibles avec ceux de nos recherches antérieures [1], nous avons utilisé le coefficient de corrélation inter-spectrale (R) comme mesure de la similarité des spectres. Un coefficient de corrélation a donc été calculé pour chaque comparaison de deux spectres entre eux.

Résultats

Pour chaque locuteur, on procède à la comparaison de chacune de ses productions francophones

<table>
<thead>
<tr>
<th>R</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
</tr>
</thead>
<tbody>
<tr>
<td>.78</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.79</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.80</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.81</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.82</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.83</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.84</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.85</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.86</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.87</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.88</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.89</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.90</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.91</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.92</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.93</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.94</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.95</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.96</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Tableau 1 : fréquences absolues des coefficients de corrélation intra-locuteur inter-langue, par locuteur.

à chacune de ses productions néerlandaises. Ce traitement produit 1000 coefficients de corrélation inter-langue intra-locuteur (10 locuteurs x 10 productions françaises x 10 productions néerlandaises). Les distributions de ces coefficients de corrélation sont présentées au tableau 1, sujet par sujet et pour l'ensemble de ceux-ci.

<table>
<thead>
<tr>
<th>Sujets</th>
<th>Types de comparaison</th>
<th>m gén.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F/F</td>
<td>NL/NL</td>
</tr>
<tr>
<td></td>
<td>m</td>
<td>sd</td>
</tr>
<tr>
<td>S1</td>
<td>.935</td>
<td>.009</td>
</tr>
<tr>
<td>S2</td>
<td>.932</td>
<td>.010</td>
</tr>
<tr>
<td>S3</td>
<td>.931</td>
<td>.015</td>
</tr>
<tr>
<td>S4</td>
<td>.925</td>
<td>.014</td>
</tr>
<tr>
<td>S5</td>
<td>.958</td>
<td>.008</td>
</tr>
<tr>
<td>S6</td>
<td>.949</td>
<td>.014</td>
</tr>
<tr>
<td>S7</td>
<td>.911</td>
<td>.019</td>
</tr>
<tr>
<td>S8</td>
<td>.944</td>
<td>.014</td>
</tr>
<tr>
<td>S9</td>
<td>.942</td>
<td>.018</td>
</tr>
<tr>
<td>S10</td>
<td>.897</td>
<td>.022</td>
</tr>
<tr>
<td></td>
<td>m gén.</td>
<td>.932</td>
</tr>
</tbody>
</table>

Tableau 2 : Moyennes (m) et écarts types (sd) par sujet des coefficients de corrélation émanant de comparaisons intra-sujet inter-et intra-langue des textes français (F) et néerlandais (NL).

Afin d'assurer la comparaison de ces données avec nos résultats antérieurs [1], nous avons fait...
figurer au tableau 2 des résumés statistiques des comparaisons intra-langue, intra-locuteur des productions françaises ("F"). Rappelons que pour chaque locuteur, une comparaison interspectrale fut réalisée pour chacune de ses 45 paires non redondantes de spectres différents (soit la moitié d'une matrice 10 x 10 préalablement ajustée de sa diagonale).

Nous avons en outre, répliqué un traitement identique sur les productions néerlandaises ("NL") des locuteurs. Les résultats de ce traitement sont présentés de manière résumée au tableau 2. Celui-ci comporte en outre un résumé statistique ("F/NL") des données du tableau 1 compatible avec les deux précédents.

**DISCUSSION**


Il apparaît en outre que pour chaque sujet considéré isolément, ses corrélations moyennes intra-sujet intra-langue provenant respectivement des corpus français et néerlandais (colonnes "F/F" et "NL/NL" du tableau 2) présentent peu de différences. Une analyse de variance à deux dimensions croisées (plan mixte; sujets : dimension aléatoire; conditions F/F et NL/NL : dimension fixe) a confirmé cette observation (F = 1.7; p = .66).

Par contre, il est très clair que lors des comparaisons inter-langue, tous les sujets présentent des corrélations moyennes systématiquement plus faibles que lors des comparaisons intra-langue : dans le tableau 2, chaque moyenne de la colonne "F/NL" est plus petite que les deux autres moyennes de la ligne où elle se trouve. Cette observation a été confirmée par une analyse de variance à deux dimensions croisées appliquée à un plan ne différant de celui de l’analyse précédente que par l’adjonction d’un niveau supplémentaire (la condition F/NL) à la dimension fixe. Ce simple ajout a provoqué l’émergence d’un effet "conditions" significatif (F = 37.4; p < 10^-4). Étant donné l’absence d’effet "conditions" avérée par l’analyse précédente, il est légitime de conclure que ce nouvel effet est imputable à une variation inter-langue systématique. À noter également un important effet "sujets", qui confirme les observations antérieures (F = 120.9; p < 10^-4) ainsi qu’un important effet d’interaction (F = 30.57; p < 10^-4).

**CONCLUSION**

Notre analyse montre clairement que le changement de langue induit des variations systématiques dans le SMLT : les ressemblances intra-locuteur inter-langue entre SMLT provenant du français et SMLT provenant du néerlandais sont moins importantes que les ressemblances intra-langue de SMLT provenant de l’une ou l’autre des deux langues.

Néanmoins, il convient de noter que, même si les corrélations moyennes intra-sujet inter-langue sont assez faibles (.887, en moyenne), elles sont toujours fortement élevées par rapport aux corrélations inter-sujet intra-langue (.805, en moyenne, pour le français, par exemple [1]). On peut donc conclure que l’allégnation du SMLT provoquée par le changement de langue est moins importante, en moyenne, que celle imputable aux variations inter-sujet : quand un locuteur produit deux messages, l’un en néerlandais et l’autre en français, il y a en moyenne plus de ressemblances entre ceux-ci qu’entre l’un ou l’autre et n’importe quel autre message produit dans la même langue par un autre locuteur. Bien qu’ayant établi l’existence d’effets "langues" systématiques, nous rapprochons donc nos conclusions de celles de Tosi [3], qui conclut à l’"invariance relative" du SMLT.

En outre, l’absence de différences systématiques entre corrélations moyennes intra-sujet inter-langue entre le français et le néerlandais semble aller à l’encontre des conclusions de Majewski et Hollien [2]. Notons cependant que pour rendre les expériences parfaitement comparables, il conviendrait de tenir compte de l’ensemble des distributions inter-sujet dans les deux langues.

Finalement, l’effet inter-langue mis en évidence apporte une preuve supplémentaire de la dépendance du SMLT vis-à-vis du contenu du signal de parole, lui-même diversement influencé par différentes langues. Il convient donc de rester prudent dans la généralisation de nos conclusions à d’autres langues que le néerlandais et le français.

**REFERENCES**


EXPERIMENTAL ANALYSIS ON VARIABILITY OF SPEECH SPECTRUM OF A SPEAKER OVER TIME

K. Nagata, H. Murakami and K. Saikawa

Faculty of Engineering, Hokkaido University, Sapporo, 060 Japan

INTRODUCTION

Words uttered at different times over long periods by the same speaker, do not always have the same spectrum time patterns. This variability is the main cause for increases in error rates in automatic verification of speakers [1],[2],[3].

The main cause of this variability has been assumed to arise from variations at the glottal source. To investigate the cause of the variability, vowel sounds where spectra varied were analyzed into transfer-functions of vocal tracts and glottal waveforms using short-term LPC analysis [4]. The results of the experiments indicate that the variability of the transfer-functions is dominant.

A STUDY ON THE SHORT-TERM LPC ANALYSIS

A study on the short-term LPC analysis was made with several synthetic vowel sounds. Three kinds of transfer-functions of vocal tract filters and seven kinds of glottal sources were used for the synthetic vowel sounds. Fig.1 shows an example of the transfer-functions of vocal tract filters which has 3 poles. Fig.2 shows an example of the glottal waveforms, which was extracted from a spoken vowel sound. To approximate the vocal radiation factor, each synthetic vowel sound was computed with time difference calculations from an output signal of each vocal tract filter excited by each glottal source. Short-term LPC analysis using the covariance method was applied to synthetic vowel sounds to estimate transfer-functions of vocal tracts and glottal waveforms. The distance, $d^2$, between the transfer-function of the vocal tract filter and the transfer-function estimated from the synthetic vowel sound by the analysis was measured:

$$d^2 = \sum \frac{(C_k - C_{k0})^2}{C_k}$$ (1)

ORDER OF VOCAL TRACT FILTER : $p = 6$

SPECTRAL ENVELOPE [DB]

FREQUENCY [KHz]

1.0  2.0  3.0  4.0

Fig.1. An example of transfer-functions of vocal tract filters for synthetic vowel sounds.

C_k : the kth order cepstrum coefficient estimated from the synthetic vowel.
C_k0 : the kth order cepstrum coefficient of the transfer-function of the filter.

Short-term LPC analysis with 20 sampling points analysis segments was successively applied by shifting one sampling point along the time axis. Fig.3 shows an example of the relations between $d^2$ and the first sampling points of the analysis segments (---). In the figure, $d^2$ for the glottal source spectrum and the estimated glottal source spectrum are also plotted (---). The horizontal axis represents the first sampling point of the analysis segments; 0 is the sampling time when the vocal tract filter is excited by the main glottal excitation, and corresponds to the end point of the open time of the glottis. Here, the transfer-function of the filter shown in Fig.1 and the glottal wave which has the zero amplitude time segment, were used. An LPC analysis of the 8th order was applied to the time differential signal of the synthetic vowel signal.

Fig.4 shows another example for a synthetic vowel sound which has a vocal tract filter of 4 poles and the glottal source shown in Fig.2.

Fig.2. An example of glottal waveforms for synthetic vowel sounds.

Fig.3. $d^2$ versus location of the first sampling point of analysis segments.

Fig.4. $d^2$ versus location of the first sampling point of analysis segments.
These figures allow the conclusion that the first sampling point of the analysis segment must be located just after the main glottal excitation.

Fig. 5 shows prediction residuals, estimated glottal waveforms, and estimated transfer-functions derived by the short-term LPC analysis successively applied to a spoken vowel sound /a/, by shifting one sampling point. The left end points of the residuals are shown as the first sampling points of the analysis segments, therefore the estimated glottal waveform and the estimated transfer-function marked with circles, are derived from the analysis segment with the first sampling point located just after the main glottal excitation.

The glottal waveform and the transfer-function marked with circles are finally estimated as the spoken vowel sound.

VARIABILITY OF SPEECH SPECTRUM OF A SPEAKER

To analyze the variability of speech spectra of a speaker over time, utterances of Japanese spoken number digits by males were used.

Sample utterances were recorded 7 times through one year. Utterances of 4 of the males with relatively large spectra variation of vowel sounds were analyzed further. Reference utterances were selected from the utterances at the first recording time. Segments of test vowel sounds /a/, /e/, /o/ which correspond to stationary vowel segments of reference utterances by DP time matching, were extracted from the test sample utterances.

24 test samples uttered at the same time with reference utterances and 28 test samples uttered at other times, were selected and analyzed by the 12th order short-term LPC analysis mentioned above.

A transfer-function and a glottal waveform estimated from a test vowel sound were compared with the reference vowel sound, and the similarity and \( d^2 \) between them were observed.

Table I shows classification of these similarities and \( d^2 \) of each group. It is seen that the variability of the transfer-function is larger than that of the glottal source, and the variability over time is larger than that for utterances uttered at the same time.

CONCLUSION

The short-term LPC analysis was investigated to estimate transfer-functions of the vocal tract and glottal waveforms from vowel sounds. Estimates are not perfect, but are considered adequate for these experiments.

52 vowel segments of utterances uttered over one year were analyzed by the LPC analysis. And it was found that the variability of transfer-functions of the vocal tract is generally dominant.

TABLE I SPECTRAL VARIATIONS OF VOWEL SEGMENTS OF TEST UTTERANCES WITH VOWEL SEGMENTS OF REFERENCE UTTERANCES OF EACH SPEAKER

<table>
<thead>
<tr>
<th>Number of Samples</th>
<th>( d^2 )</th>
<th>Number of Samples</th>
<th>( d^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varied 5</td>
<td>Mean 0.415</td>
<td>22 Mean 0.945</td>
<td></td>
</tr>
<tr>
<td>Transfer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varied 13</td>
<td>Max 1.022</td>
<td>4 Max 1.997</td>
<td></td>
</tr>
<tr>
<td>Function Slightly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost 6</td>
<td>Min 0.079</td>
<td>2 Min 0.393</td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Varied 2</td>
<td>Mean 0.164</td>
<td>5 Mean 0.101 (0.160)</td>
<td></td>
</tr>
<tr>
<td>Glottal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max 0.418</td>
<td></td>
<td>10 Max 0.537</td>
<td></td>
</tr>
<tr>
<td>Source Slightly</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost 5</td>
<td>Min 0.009</td>
<td>13 Min 0.035</td>
<td></td>
</tr>
<tr>
<td>Same</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

COMPARISON OF SPECTRAL SIMILARITY INDICES FOR SPEAKER RECOGNITION

B. Harsoghmien and A. Landercy

Département de Phonétique et Psycho-acoustique, Universite de l'Etat à Mons, 7000 Mons, Belgium.

INTRODUCTION

Long-term speech spectra (LTS) have been used with success in several speaker-recognition experiments [1-4]. Such experiments require the use of metrics allowing to associate with each pair of spectra one value indicating the degree of (dis)similarity. Indices of similarity vary quite a lot among papers and little is known about their respective discriminating powers, as most of the time, different indices are used in different experimental conditions. Moreover, the rare experiments allowing such investigations [4, 5] have dealt with spectra defined on a relatively restricted number of frequency channels and/or obtained by means of constant relative bandwidth analyser.

In this paper, we will study the relative discriminating powers of similarity indices applied to high-resolution, constant-bandwidth spectra. Two classical indices will be studied (i.e., the Euclidean distance, ED, and the cross-correlation coefficient, R), as well as two original ones (i.e., the chi-square index, C2, and the standard deviation of the spectral differences distribution, SD).

SIMILARITY INDICES

The Euclidean distance can be used as a measure of the dissimilarity of two spectra considered as two points in a multi-dimensional space. In our study, the 400 intensity values for each spectrum were considered as its coordinates in a 400-dimensional space. Let us consider S and S', two spectra defined by their 400 intensity values S and S'. The Euclidean distance between them was defined as:

$$ED_{ii'} = \sqrt{\sum_{i=1}^{400} (S_i - S_{i'})^2} \quad (1)$$

The cross-correlation coefficient varies from 0 (total dissimilarity) to 1 (perfect similarity) in absolute values. It was computed, in our study, as follows:

$$R_{ii'} = \frac{1}{400} \sum_{i=1}^{400} \frac{(S_i - M_S)(S_{i'} - M_{S'})}{\sigma_S \sigma_{S'}} \quad (2)$$

Where $M_S$ and $M_{S'}$ are the means for respectively all intensity values $S_i$ and $S_{i'}$, and $\sigma_S$ and $\sigma_{S'}$ are the corresponding standard deviations.

The chi-square statistic is usually used to determine whether two (or more) distributions are similar or not. Our chi-square index was drawn from this basic statistical tool. It was expected to measure to what extent energy is distributed in the same way across the 400 frequency channels of two spectra. $C_2$ was defined as:

$$C_{2ii'} = \frac{1}{400} \sum_{i=1}^{400} \frac{(S_i - S_i')(S_i - S_{i'})}{T_{S_i} T_{S_{i'}}} \quad (3)$$

Where $T_{S_i} = \sum_{j=1}^{400} S_j \cdot S_j'$ and $T_{S_{i'}} = \sum_{j=1}^{400} S_j' \cdot S_j$. The standard deviation of the differences distributions measures the variability of the differences $S_i - S_{i'}$, across the whole frequency range under investigation. High SD indicate important dissimilarities between spectra, and conversely. SD was defined as:

$$SD_{ii'} = \sqrt{\frac{1}{400} \sum_{i=1}^{400} [S_i - S_{i'} - MD]^2} \quad (4)$$

Where MD is the mean for the $S_i - S_{i'}$ differences.

R and SD are insensitive to variations in the overall levels of the compared spectra. On the contrary, spectral comparisons by means of ED or by means of $C_2$ require a prior intensity-normalisation. This one was performed during a pre-processing stage where the overall levels of the LTS were set at an arbitrary total level of 100 dB.

EXPERIMENTATION

Experimental setting

The speakers were 10 French-speaking male subjects, between 19 and 21 years old. Each of them was asked to utter a phonetically balanced French text ten times in succession. The text was about 18 sec. long.

The utterances were recorded on a Nagra IV S recorder, by means of a KM 84 Neumann microphone. The recording sessions took place in a quiet sound-proof room; the subjects were sitting in front of the microphone, placed at a constant 40 cm distance from their lips.

The recordings were afterwards analysed by means of a Bruel Kjaer 2033 400 channels FFT analyser (NB 2033). Its sampling frequency was set at 12.6 KHz, determining a constant 12.5 Hz resolution across the whole 0-5 KHz range of analysis. A linear averaging process was moreover used in order to compute one LTS for each utterance.

The 100 (10 subjects x 10 utterances) so-obtained LTS were then transmitted from the analyser to an Apple II computer via a GPIB interface card, for further computations.

Comparison procedure

Inter- as well as intra-speaker comparisons were performed. For each of the 10 speakers, one comparison was performed for each possible non-redundant pair of his 10 LTS, i.e., 45 comparisons by subject, thus 450 intra-speaker comparisons for the whole sample. Symmetrically, for each possible non-redundant pair of different speakers (i.e. 45 pairs of speakers), 100 inter-speaker comparisons were achieved, i.e., 4500 inter-speaker comparisons for the whole sample.

For each comparison of two spectra, the four similarity indices under investigation were computed.
RESULTS

For each similarity index, 450 intra-speaker values and 4500 inter-speaker values were computed. These values constituted the inter- and intra-speaker distributions we used to evaluate the discriminatory ability of our indices. For each index, a series of values selected across its entire range of variation was successively considered as rejection thresholds for a recognition task. The corresponding false alarm-and correct recognition rates were drawn from the observed distributions: four ROC curves (one by index) were drawn from this statistical processing. They are presented in figure 1.

DISCUSSION

Our experimental results clearly show that the four indices tested have good discriminatory abilities. The cross-correlation coefficient, nevertheless exhibit results weaker than those of ED, SD and C2, which appear to be almost equivalent (with slightly better performances for ED). This conclusion is rather contradictory to Zalewski at al.'s finding who reported identification rates slightly better with R than with ED.

Finally, the standard deviation of the spectral differences distribution (SD) appears as a promising index. On the one hand, its discriminatory ability is the nearest to the one of DE; on the other, it does not require any intensity - normalization. It can thus be considered as the best compromise between processing cost and discriminating capacity, so far as our four indices are concerned.

REFERENCES


Figure 1: Relative Operating Characteristics curves for the four indices tested.

Swets [6] has shown that the proportion of the area of the entire ROC space that lies beneath the ROC curve is a distribution-free measure of sensitivity. This property allowed us to perform a ranking of our indices. The worst index is obviously R, for which the surface under the ROC curve is sensibly smaller than for others. The most discriminant index appears to be DX, whereas SD and C2 are respectively slightly less discriminant.
AUTOMATIC RECOGNITION OF SPOKEN SENTENCES USING A DEMISYLLABLE-BASED DYNAMIC PROGRAMMING ALGORITHM

G. Ruske and W. Weigel


1. INTRODUCTION

The continuous speech recognition system presented in this paper aims to decode spoken sentences as a series of words. The word level is generally important because the words are the fundamental units for syntactic and semantic analysis. Since the investigations described below focus on the acoustic-phonetic problems of continuous speech, the system works in these experiments without syntactic and semantic constraints.

For large vocabularies it is advantageous to use smaller units from which all words can be built up. The presented system starts from an explicit segmentation of the speech signal into demisyllable segments. Each demisyllable contains a consonant cluster in syllable initial or final position, and the vowel from the syllable nucleus. Consonant clusters and vowels are assigned phonetic symbols by a classifier evaluating the loudness spectra within the demisyllable segments. Each word of the vocabulary is represented by a phonetic word model containing the variations in pronunciation as well as possible segmentation errors. The word models are realized as graphs with directed arcs, which can be processed very efficiently using a 1-stage Dynamic Programming (DP) algorithm. This algorithm performs word and sentence recognition in one step in order to find that path through a series of models which gives the best match with the phonetic symbols provided from the classification stage.

2. SYLLABIC SEGMENTATION AND CLASSIFICATION

Segmentation is based on preprocessing of the speech signal using a model of loudness sensation; these methods have been reported in other papers /1,2,3/ and will be summarized here only very briefly. The loudness model consists of a critical-band-rate filter bank with 22 channels (50 Hz - 8.5 kHz) each representing a loudness component. All 22 components constitute a so-called loudness spectrum which is sampled every 10 ms. The syllable nuclei are localized from the peaks of a smoothed loudness function as well as from spectral information. Syllable boundaries are indicated by loudness minima. The demisyllable segments span the range from a syllable boundary to the center of the syllable nucleus.

Classification of demisyllable segments is performed by using a nearest-neighbour classifier which operates on the basis of normalized templates for initial consonant clusters, final consonant clusters, and vowels or diphongs; details are reported in /1,2,3/. The classifier yields phonetic symbols which are abbreviated in this paper as follows:

I = Initial consonant cluster,
V = Vowel or diphong,
F = Final consonant cluster.

It is characteristic for this approach that the symbols are provided always in the fixed order I-V-F per syllable.

3. LEXICAL REPRESENTATION

The word models have to capture the variations in pronunciation as well as the possible syllabic segmentation errors from the preceding segmentation and classification step. Experience with this system has shown that in practice never two consecutive insertions or omissions of a single syllable occur. Therefore the models are provided with i-syllable skips which are realized as directed arcs in a graph representation, see fig. 1. The arcs bear the phonetic symbols whereas the nodes have no special meaning here. In this basic model an omission of a syllable is possible at each position (at each node in the graph) by using the corresponding skip arc.

Fig. 1. Basic word model with i-syllable skips (I-initial consonant cluster, V-vowel, F-final consonant cluster).

An insertion of a syllable has to be realized in the horizontal main path which for its part can be passed over by a skip describing the "normal" pronunciation. That means, the model is able to cover the following cases:

- the main path represents the normal pronunciation and can be skipped syllable-wise in the case of a syllable omission;
- the main path represents an insertion of a syllable whereas a skip over this syllable means normal pronunciation.

The skips principally allow the replacement of a complete syllable by an initial consonant cluster, a final consonant cluster, or a vowel; the appropriate usage depends on the kind of segmentation error to be handled. For instance, the German word "legen" (/lejgn/) can be reduced to /lejn/ which is expressed by a skip over the second syllable replacing it with the final consonant cluster /n/. In the same way replacements can be performed by use of an initial consonant cluster or a vowel.

Each arc of the model in fig. 1. is given a weight thus accounting for the fact that some pronunciations occur more often than others.

The word models have been constructed automatically by a training procedure applied to a training set of classified sentences. In these contexts it is only necessary to mark the word boundaries by hand whereas the training of the word models itself operates unsupervised.

4. SENTENCE RECOGNITION

Sentence recognition is based on the well-known 1-stage Dynamic Programming (DP) algorithm for connected word recognition /4/. In contrast to the classical usage of spectral frames as a parametric representation of speech this algorithm was modified here to enable the processing of the demisyllable word models described above. Thus as indicated in fig. 2 the elements along the horizontal axis of the distance matrix are the syllable classification results (symbols) delivered by the classifier. The different word models of the lexicon are represented by four columns along the vertical axis from right to left: Two containing the symbols for the main path in the word model and the skip arcs and two containing the weights for these arcs according to the general model in fig. 1. Before explaining the processing of the word models by the algorithm in detail it is necessary to describe some modifications of the usual DP parameters.
Fig. 2. 1-stage DP algorithm for sentence recognition operating on the symbol level.

Firstly, a suitable distance measurement has to be defined which in our case requires a measure of similarity between a symbol of the word model denoted by \( y \) and a symbol of the unknown utterance denoted by \( x \). It was pointed out in former investigations that the \( \alpha \)-posteriori probabilities are suitable for this purpose. Due to the nature of \( \alpha \)-posteriori probabilities the term \( p(y|x) \) means the probability of having really spoken the class \( y \) for the case that the classifier delivered the decision \( x \). Because of the accumulation of distances by the DP algorithm we use the logarithm of the probabilities to get an estimation of the joint probability along one alignment path. So the local distance measurement for one point in the matrix is calculated as:

\[
d(x,y) = \log p(y|x)
\]

with \( x \): classification result, \( y \): symbol of the word model.

The second important parameter of the DP is the choice of the transition rule, which defines the allowed transitions for the paths through the distance matrix.

![Transition diagram](image)

The transition rule used is illustrated in fig. 3; the rule is derived from the two types of arcs in the word models, namely the horizontal one and the skip arc.

Fig. 3. Transition diagram.

In fig. 2 marked by a square with each possible start node of a word model (marked by a circle) at each instant in the utterance. Combining both rules one achieves a matching of a series of word models onto the input symbol sequence. Backtracking the best path yields that word chain with the optimal similarity to the spoken input.

5. Experimental Results

In order to examine the recognition accuracy of the described methods we used a corpus of 16 different German sentences containing 75 different words and 77 words in all. These words represent a subset of the 1001 most frequent German words. The sentences were spoken by a male speaker six times. Three versions of the sentences built up the learning set for an algorithm which generates the word models automatically by establishing the arcs of the word models with the best symbolic representation. Since the speech material was relatively small, it seemed reasonable to give equal values to all the weights of the arcs in all word models. Using the other three versions (48 sentences) as test material we obtained the results shown in Table 1.

<table>
<thead>
<tr>
<th>Recognition score</th>
<th>Errors (words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sentences</td>
<td>words</td>
</tr>
<tr>
<td>21</td>
<td>222</td>
</tr>
</tbody>
</table>

Tab. 1. Recognition results from 48 test sentences.

It is important to point out that these results are achieved without any grammatical or semantic constraints. Thus an achieved word recognition score of about 85% can be regarded as quite encouraging.

Of course, the results also depend on the quality of the word models as well as on the \( \alpha \)-posteriori probabilities expressing the possible confusions. For this purpose an algorithm for automatic generation of suitable word models has been developed which will be published in a subsequent paper.

References:

A PITCH SYNCHRONOUS WEIGHTED LINEAR PREDICTION ANALYSIS OF SPEECH

M. Yanagida and O. Kakusho

The Institute of Scientific and Industrial Research, Osaka University
Mihogaoka S-1, Ibaraki, Osaka 567 Japan

Proposed here is a means to speed up the computation process of the Weighted Linear Prediction (WLP) analysis method developed by the authors. The WLP attains outstanding accuracy compared to the ordinary linear prediction analysis, but it requires lots of preliminary analysis and inverse filtering for each analysis frame. The proposed method shifts the analysis frame pitch synchronously aiming at reduction of computational loads. The basic idea is that a common weight sequence may be employed in a series of analysis frames since the prediction residual seems to repeat almost the same patterns pitch synchronously. In the proposed Pitch Synchronous Weighted Linear Prediction (PSLWP), a single weight pattern is repeatedly employed pitch synchronously in a series of analysis frames instead of calculating weight patterns frame by frame. The simulation results on synthetic vowels show that the proposed PSLWP attains almost the same accuracy as the proto-type WLP reducing the computational time down to 30%.

INTRODUCTION

The authors have developed a Sample Selective Linear Prediction (SSLP) method[1] and a Weighted Linear Prediction (WLP) method[2] as its generalized version aiming at improving the accuracy of the least-squares estimate for the predictor coefficients by cutting off or reducing the contribution of prediction equations that contain large prediction errors. Both SSLP and WLP, however, require a lot of computation time as they need preliminary analysis and inverse filtering for each analysis frame, though they have much improved the accuracy in estimating formant frequencies. Particularly WLP needs calculation for weighting after inverse filtering for every frames, though it gives better estimation results than SSLP. This paper describes a means to speed up the computation process of WLP without losing accuracy. Proposed is pitch synchronization of frame shifting intending reduction of computational loads for preliminary processing for each analysis frame. By making frame shift pitch synchronous, those procedures can be abbreviated except for the first frame of a series of analyses, since the residual sequence is supposed to repeat pitch synchronously for voiced segments uttered by a speaker even for different phonemes and even if its pitch varies.

WEIGHTED LINEAR PREDICTION

The linear predictive coding of speech is modeled that the n-th sample \(y_n\) is approximated by a linear combination of past \(p\) samples as

\[
y_n = \sum_{i=1}^{p} a_i y_{n-i}
\]

where \(a_i\) denotes the \(i\)-th predictor coefficient. The linear prediction analysis, in particular the covariance method, means to get a set of \(a_i\)'s that minimizes the squared error. Eq.(1) can be written

in matrix form as follows

\[
y \dagger y \alpha
\]

where

\[
y = \begin{bmatrix}
y_n \\
y_{n+1} \\
\vdots \\
y_{n+p-1}
\end{bmatrix}
\]

\[
y = \begin{bmatrix}
y_{n+1} \\
y_{n+2} \\
\vdots \\
y_{n+p}
\end{bmatrix}
\]

\[
\alpha = \begin{bmatrix}
a_1 \\
a_2 \\
\vdots \\
a_p
\end{bmatrix}
\]

The least-squares estimate for \(\alpha\) is given as the solution of the following Yule-Walker equation obtained by multiplying \(y^\dagger\) to eq.(2) from the left as

\[
y^\dagger y \alpha = y^\dagger y \alpha
\]

Then,

\[
\hat{\alpha} = (y^\dagger y)^{-1} y^\dagger y = y^\dagger y
\]

where \(\dagger\) denotes the generalized inverse of the least-squares type.

Putting weight \(\omega_n\) on both the sides of eq.(1) we can express the predictive equation in matrix form as

\[
W \hat{y} = W \hat{y} \alpha
\]

where

\[
W = \text{diag} \{ \omega_n, \omega_{n+1}, \ldots, \omega_{n+p-1} \}
\]

The least-squares solution \(\hat{\alpha}\) for this case is expressed as

\[
\hat{\alpha} = (W^\dagger W)^{-1} W^\dagger W
\]

This is the formal expression of the Weighted Linear Prediction (WLP). Eq.(4) is the key equation of the WLP as \(\hat{\alpha}\) can be obtained by back substitution procedure after applying the Given's reduction[3] to eq.(4) directly or by employing a dyadic function of AP[4] as

\[
\hat{\alpha} = (W^\dagger W)^{-1} (W^\dagger y)
\]

In order to obtain accurate estimates for the predictor coefficients, the weight should be made small for prediction equations containing large prediction error so the weight may effect to reduce the contribution of such equations. Therefore, the weight \(\omega_n\) for prediction equation for \(y_n\) is made as a function of the prediction error \(e_n\) employing the exponential part of the normal distribution function as follows;

\[
\omega_n = \omega(e_n) = \exp \left( -\frac{1}{2} \left( \frac{e_n}{\sigma} \right)^2 \right)
\]

where \(\sigma\) is empirically determined to be 0.5 normalized by the maximum value of \(e_n\).

The authors[2] have confirmed that the WLP gives more accurate estimates for the formant frequencies of synthetic sounds than the ordinary covariance method.

PITCH SYNCHRONOUS WEIGHTED LINEAR PREDICTION

The proto-type WLP, however, has difficulty that preliminary analysis and inverse filtering are required before the final analysis for each analysis frame in order to calculate the weight for prediction equations in the analysis frame. However, most of these preparatory procedures might be omitted by assuming that the residual signal repeats similar patterns every pitch interval for a speaker.
even if its fundamental frequency varies.

If the residual signal can be assumed to repeat the same pattern every pitch period, a common weight pattern can be employed for every analysis frame, shifted synchronously to pitch. Therefore, the calculation for determining the weight is required only for the first analysis frame. Thus, the calculation process of the WLP is drastically reduced. We call this method a Pitch Synchronous Weighted Linear Prediction (PSWLP). The window length for PSWLP is preferably made one pitch period long so as the analysis is independent of the window position. In PSWLP the correspondence between the sample points on the speech signal and that on the weight pattern has a severe significance, then the shifting interval should be strictly synchronized to pitch interval. For stationary synthetic sounds, frame shifting can be controlled by detected pitch as the pitch detection algorithm does not fail on stationary synthetic sounds.

However, for real speech sounds, there may be some amount of error in pitch extraction and the residual signal is not exact periodic repetition of one pitch residual signal. So, simple one-pitch shifting may cause mismatch in correspondence between sample points and weight values, and the shifting interval for the analysis of real speech should be synchronized to local maximal point on voiced parts of speech waves.

The processing algorithm of the PSWLP is depicted in Fig.1.

![Image of processing scheme of PSWLP](image)

**Fig.1** The processing scheme of the PSWLP

**ACCURACY IN FORMANT FREQUENCY EXTRACTION**

In order to confirm the performance of the PSWLP, accuracy comparison in estimating formant frequencies was made among the following methods:

- **COV(N)**: covariance method (N:window length)
- **SSLF(S)**: Sample Selective LP
- **WLP**: Weighted Linear Prediction
- **PSWLP**: Pitch Synchronous WLP

An all-pole filter of order 10 was employed to synthesize vowel-like test sounds excited by one-pitch residual sequence of 90 samples long.

Figure 2 shows the results of the comparison. The results are evaluated by estimation error(avg) averaged for 10 different window locations and for five formants over 50 synthetic sounds of different formant patterns. The analysis order is 12 on differenced data with pre-emphasis factor 0.03. The abscissa is 0 for WLP and PSWLP and B for SSLF. From Fig.2, it can be seen that the proposed PSWLP retains the same accuracy level as the WLP for broad range of a although the window length is as short as one pitch period.

![Image of comparison errors on stationary synthetic sounds](image)

**Fig.2** Comparison in estimation errors on stationary synthetic sounds.

**REDUCTION OF REQUIRED PROCESSING TIME**

The processing time for the above-mentioned methods are compared in Table 1, where the average time required for processing one frame are shown. The processing time required for PSWLP is reduced down to 30% of that required for WLP.

**Table 1** Comparison of the processing time required. (on UX MV/8000 11)

<table>
<thead>
<tr>
<th>Analysis Method</th>
<th>Time Required(sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariance method</td>
<td>0.04</td>
</tr>
<tr>
<td>Givens’ reduction</td>
<td>0.09</td>
</tr>
<tr>
<td>WLP(2)</td>
<td>1.22</td>
</tr>
<tr>
<td>PSWLP</td>
<td>0.36</td>
</tr>
</tbody>
</table>

**CONCLUSIONS**

The PSWLP, an efficient algorithm for WLP, is proposed introducing pitch synchronisation in frame shifting. It eliminates the weight calculation process of succeeding frames and retains the same analysis accuracy reducing the required processing time down to 30% of WLP.

Acknowledgment: This work has been partly supported by Grant-in-Aid for Scientific Research for 1984-1985 (No.59550039) from the Ministry of Education, Science and Culture of Japan.

The authors express their sincere thanks to Mr. M. Tsunoda for his help in computer simulation.

**REFERENCES**


OPTIMIZING PITCH PERIOD MARKERS PRIOR TO EXTRACTING FEATURES FROM ISOLATED VOWELS

N.D. Cox, M.D. Morrison and M. Rito

Voice Lab, Vancouver General Hospital, 4th Floor
Willow Pavilion, 805 West 12th Avenue, Vancouver, B.C., V5Z 1M9

Feature extraction from isolated vowel samples is frequently preceded by a pitch-period marking process in which the 'startpoint' of each pitch period is located by marking the position of each occurrence of an arbitrarily selected waveform event. The startpoints optimally correspond to a fixed point in the vocal cord vibratory cycle. Unfortunately, the relative position of simply defined waveform events often varies from pitch period to pitch period and intersample variability makes it difficult to define a simple nonambiguous startpoint event for use in all vowel samples.

The purpose of this paper is to discuss the use of cross-correlation for 'optimizing' pitch period startpoint markers. Given a digitized sample with previously marked startpoints, the algorithm adjusts the startpoints so as to maximize the correlation between the waveforms in the region of each startpoint. The method does not rely on the relative position of any particular waveform event and may be applied to any set of startpoints.

The algorithm was evaluated using synthetic /a/ vowels containing random perturbation. It was found that the optimized startpoints were synchronized with the driving function to within 1 sample point for all but the most severely perturbed waveforms. The algorithm effectively compensated for the high sensitivity of the Harmonics to Noise Ratio (1) to pitch period misalignment.

The Pitch Period Startpoint Optimization Algorithm

The pitch period startpoint optimization algorithm is based on interpolation of a correlation function. Similar algorithms have been used in other areas (2). A short waveform segment following the first startpoint is used as a reference segment. The reference segment is convolved with the region surrounding the second startpoint and the correlation coefficient is computed for a number of offsets. Parabolic interpolation (2) is used to locate the correlation coefficient function peak and the startpoint marker is adjusted accordingly. The square of the correlation coefficient multiplied by its sign was used so as to avoid square root computations. The procedure is similarly repeated for each startpoint.

Text Waveform Synthesis

Nine 10kHz and nine 20kHz /a/ vowels each with equal percentages of jitter, shimer and noise and a 12kHz fundamental frequency were synthesized using a filtered impulse train. Waveforms were generated at 40kHz and downsamped prior to storage so as to obtain noninteger pitch period startpoints. Random jitter, shimmer and noise perturbations were implemented as percentages of the impulses spacing, impulse size and unperturbed output amplitude respectively. Perturbations ranged from 0% to 50%. Startpoint markers accurate to 3 decimal places were synchronized with the driving impulses and offset to be aligned with the waveform transition proceeding the first large pitch period oscillation.

Parameter Selection

The optimization algorithm was applied with various segment sizes to 50 consecutive pitch periods in each test file to identify a value which minimizes marking error variance. The segment size is the length of the reference segment and the marking error is the difference between optimized startpoints and the startpoints stored when the data was synthesized. Marking error variance is a measure of the degree of synchrony between the optimized startpoints and the vowel driving function. In general, variance decreased as the segment size increased up to a point at which no improvement could be made. The optimum value varied with perturbation level but a segment size of 2 msec performed well at all levels.

A number of preprocessing techniques have been applied in conjunction with correlation-based pitch detection algorithms (4). High frequency preemphasis, moving averaging, and center clipping were evaluated here. Center clipping proved to be detrimental. Moving averaging alone was also detrimental but a performance improvement was obtained when it was combined with high frequency preemphasis. The best overall performance for 10kHz data was obtained with 334 preemphasis and a 2 point moving average. The best 20kHz setting was 70% preemphasis and a 3 point moving average.

The range over which correlation coefficients are computed directly determines the size of marking error that can be isolated. If the range is too large, the range may result in pitch period oscillations being confused as startpoints. The range should therefore be less than the period of the first formant. A range of 1 msec was selected.

Performance

A plot of marking error variance as a function of perturbation for the above parameter selections can be found in Figure 1. The error variances resulting from quantization of the startpoints rounding to the nearest sample point in the 20kHz files are plotted for comparison. For 20kHz data the algorithm's error variances were less than 20kHz quantization variance for perturbation levels less than 20%. Even the 10kHz optimizations produced lower variances than 20kHz quantization for perturbation levels of 10% or less. The performance improvement with preprocessing is apparent at low and medium perturbation levels.

Relatively large error variances were observed for perturbation levels greater than 30%. However, in the worst cases all but a few startpoint estimates were accurate to within 2 sample points.

Use with the Harmonics to Noise Ratio

The Harmonics to Noise Ratio (HNR) estimates the ratio of the acoustic energy in the periodic component and the nonperiodic component of isolated vowels. HNR estimates computed with correct startpoints, quantized startpoints and optimized startpoints have been plotted as a function of perturbation in Figures 2 and 3. A 0.25 point offset was added to each quantized startpoint so that interpolations occurred in all estimates. Third order Lagrange interpolation was used.

Inspection of Figures 2 and 3 reveals the extreme sensitivity of the HNR to pitch period marker misalignment. Quantization alone resulted in a 10% underestimation of up to 10 dB in low perturbation waveforms and there was little variation below 3% perturbation in 10kHz estimates or below 1% perturbation in 20kHz estimates.
Startpoint optimization produced a large improvement in the INR estimates. All optimized estimates in perturbed waveforms were within 0.72 dB of the actual value for 10KHz data and within 0.31 dB for 20KHz data. The range over which perturbation levels could be differentiated was extended to below 1% for 10KHz data and approached 0% for 20KHz data.

At low perturbation levels, the 'actual' INR estimates for 10KHz data were less than the corresponding 20KHz estimates. The likely cause of this discrepancy is increased interpolator inaccuracy with coarser quantization. Therefore, 20KHz sample frequencies are indicated for good performance at low perturbation levels.

Conclusions

The startpoint optimization algorithm produced startpoints which were closely synchronized with the driving function of the synthetic vowels. Using a reference segment size of 2 msec and a search region size of 1 msec, algorithm's marking error variances were less than 10 usec2 for perturbation levels of less than 5% and remained less than 20KHz quantization variance for perturbation levels less than 2%. Preprocessing 20 KHz data with 70% high frequency preemphasis and a 3 point moving average and preprocessing 10 KHz data with 35% preemphasis and a 2 point moving average improved performance.

The results support the use of the pitch period startpoint optimization prior to extraction of the INR. Pitch period misalignment due to quantization severely affected the INR estimates, especially at low perturbation levels. Startpoint optimization effectively eliminated marking error sensitivity and substantially improved the INR's ability to resolve low perturbation levels.

The data indicated that the low perturbation 10KHz INR estimates were limited by the accuracy of the interpolator. It is therefore recommended that a 20KHz sample frequency be used.

These results are probably an optimistic indication of the algorithm's performance. The formant and fundamental frequency specifications were held constant in the synthesized vowels. Other issues such as a more sophisticated source model and source-tract interaction were not modeled. Nonetheless, if it can be assumed that the resonant characteristics of a vowel sample remain roughly stationary near the start of each pitch period and that existing startpoint markers are close to their correct values, the cross correlation should contribute to consistent and accurate estimation of any feature requiring precise pitch period marking.

References

This work was supported by a Science Council of British Columbia GREAT Scholarship and a National Health Research and Development Program Fellowship.
A1-8

ISOLATED WORD RECOGNITION USING HMM WITH DURATION DISTRIBUTION

M. Nishimura and M. Okochi

Science Institute, IBM Japan, Ltd.
5-19, Sanban-cho, Chiyoda-ku, Tokyo 102, JAPAN

INTRODUCTION

We have been investigating isolated word recognition by using Hidden Markov Models (HMM). In conventional HMM approach, duration at each state is represented by a self-loop transition probability and has exponential distribution. It is not suitable for representing phonetic duration information. To exploit the duration information, B.H. Juang et al. [5] used an estimated duration for adjusting the score of the conventional HMM, and M.J. Russell et al. [6] proposed an HMM which has a duration distribution function (Poisson distribution) at each state. To represent the duration explicitly and precisely, we introduced an HMM with an arbitrary duration distribution at each state, which was called a "duration model" [3], and improved the model at several points.

An HMM requires a large amount of training data to represent phonetic fluctuations. Furthermore, since the duration model has more parameters than the conventional HMM, it requires more training data and more processing time to estimate the parameters. For training the duration model by a limited amount of data, we introduced a parameter smoothing method. And to make the decoding efficient and precise, we introduced a path constraint window calculated from the state duration parameters. Since the duration model reflects the temporal structure in a word more precisely, the model is more sensitive to word boundary detection than the conventional HMM. We introduced a probability of a word boundary which is calculated as the recognition results of a silence model.

We conducted a recognition experiment with a vocabulary of similar 150 words. The recognition error rate was reduced from 9.0% (conventional HMM) to 2.8% (our method).

CONVENTIONAL HMM AND DURATION MODEL

A typical conventional HMM is shown in Fig.1(a). Let M be a model which represents a word. The probability that a state sequence $s(0)s(1)...s(T)$ of model M outputs a label sequence $o=0o_2...o_T$ is represented by

$G(P,M)=\prod_{t=0}^{T} q(s(t-1),s(t)) r(s(t-1),o_{t+1})$

where $q(i,j)$ is the state transition probability from state i to state j, $r(i,j,k)$ is the output probability of label k in transition from state i to state j.

In this model, each state has an exponential duration probability; the probability of duration d at state i is

$D(i,d)=(1-q(i,i)+q(i,i)e^{-d})$

where q(i,i) is the self-loop transition probability of state i. This is not suitable for representing the duration precisely.

For modeling a state duration with an arbitrary distribution, we introduced an HMM structure as shown in Fig.1(b). For duration di of state si, a branch with (di-1) intermediate states is assumed between state si and state s(i), and the output probabilities at each intermediate state are assumed to be common to those of the root state si. Let P=(s(i,di))(s(i,d2)...(s(i,do)) be a state sequence with duration di for state si. The probability that the state sequence P of model M outputs a label sequence $o=0o_2...o_T$ is represented by

$G(P,M)=\prod_{i=0}^{T} q(s(i,di),s(i,d2)...(s(i,do)))$

where q(s(i,di)) is the probability of duration di at state si, $r(s(i,do))$ is the output probability of label oj at state si, and $T=d_{1}+d_{2}+...+d_{o}, \sum \delta$. T

TRAINING

For estimating the parameters of the duration model, the conventional Baum-Welch algorithm is still available.

For smoothing the output probabilities obtained by an insufficient amount of training data, we used a weighting matrix which reflects the distance between labels [6]. In addition, we smoothed the duration probability q(s(i,di)) by

$q'(s(i,di))=\max(q(s(i,di)),e)$

where e is a very small positive value.

DECODING

Path Constraint Window

The Viterbi algorithm was used for decoding. The trained duration probabilities at each state were used for controlling the path constraint window. For state si, the left side path constraint window L(si) and the right side window R(si) are given by

$L(si)=-W$

$L(si)=L(si)+\min(di|q'(s(i,di))>e')$

$R(si)=-W$

$R(si)=R(si)+\max(di|q'(s(i,di))>e')$

where W is an estimated mismatch range at first state, e' (>e) is a very small positive value. The window is controlled also from the end boundary as shown in Fig.2, and a decoding path is allowed in the shaded area.

Boundary Detection

A typical conventional method of word boundary detection is a so-called double threshold method using normalized log power. This method is simple, but it is not easy to determine adequate threshold value. Environmental and phonetic fluctuations often cause serious errors.

To reduce errors of word boundary detection, a silence model with a conventional HMM structure is used together with the duration model. The silence model is trained using silence between words. In
decoding, as shown in Fig. 2, after an initial estimate of the word boundary is obtained by the double threshold method, the silence model is applied to a label sequence around the estimated boundary to get a probabilistic word start boundary. The word end boundary is defined by using the maximum probability in last state. This method is adaptable to the environmental fluctuations.

![Figure 2. The manner of decoding](image)

**EXPERIMENTS**

To clarify the recognition ability of the proposed method, we conducted a recognition experiment for a difficult task with similar 150 Japanese words such as "maiku", "miakou" and "gaikou" [7]. Two speakers recorded these words five times at the same session. Four of the five sets were used to train the models. And the remaining one set(A) and another one set(B) which was recorded two weeks later were used for decoding test.

For comparison, labeled DP matching, the conventional HMM[6] and the proposed method were applied to this confusing word recognition. As shown in Table 1, the duration model with smoothing was better than other methods. Path constraint window and probabilistic boundary detection worked well especially for data (B).

Furthermore, to clarify the ability of the probabilistic boundary detection method, we conducted an experiment using manually segmented data with erroneous word start boundary defined as shown in Fig. 3. Fig. 4 shows that, for the boundary detection error ranging from -20 to 20 frames, the proposed method was better than the best case of the usual method with a fixed boundary.

**Table 1 Average recognition error rates**

<table>
<thead>
<tr>
<th>Input data</th>
<th>S</th>
<th>F</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 Weeks later(A)</td>
<td>2 Weeks later(B)</td>
<td></td>
</tr>
<tr>
<td>Labeled DP</td>
<td>-</td>
<td>-</td>
<td>3.2%</td>
</tr>
<tr>
<td>Conventional</td>
<td>-</td>
<td>-</td>
<td>2.6</td>
</tr>
<tr>
<td>HMM</td>
<td>-</td>
<td>-</td>
<td>1.5</td>
</tr>
<tr>
<td>Duration Model</td>
<td>-</td>
<td>-</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>-</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>*</td>
<td>*</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>**</td>
<td>*</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**CONCLUSION**

In this paper, we proposed a duration model with parameter smoothing, path constraint window and probabilistic boundary detection. By using the method, the recognition error rate was reduced from 9.0% to 2.8% compared with the conventional HMM. The probabilistic boundary detection method worked well especially to reduce the recognition errors caused by the erroneous boundaries.

**REFERENCES**

LANGUE, FILTRE ET MODÈLE

P. Martin
Langues et linguistique, Université Laval
Ste-Foy, Québec C1K 7P4

En épistémologie de la linguistique, au niveau de l'articulation phonique, nous n'avons pas fait de progrès significatif depuis l'avènement de la théorie des unités discrètes, ainsi que fonctionnelle européenne avait fait faire un bon en avant en proposant la théorie du phonème et en l'opposant à la théorie du son. Les traits se trouvaient dorénavant définis non pas en termes positifs, selon les par de l'acoustique, auditif, ou articulatoire, mais en termes négatifs pour ainsi dire, selon leur capacité à sous-tendre des oppositions entre monèmes (morphèmes) différents. La phonologie, en effet, est fondée sur la pertinence communicative. Les unités, dégagées en fonction de la manière dont elles occupent dans le réseau des relations et des rapports, proportionnels ou autre, en présence. Pour une langue donnée, un phonème est ce que les mots unis à même type de son n'ont pas, dans cette langue. Du reste, le phonème peut s'articuler en paquets de traits simultanés occupant une position particulière dans la chaîne, les phonèmes, ou en supras-segments qui subissent ce-ci, les proches d'un même type de son n'ont pas, dans cette langue. Du reste, le phonème peut s'articuler en paquets de traits simultanés occupant une position particulière dans la chaîne, les phonèmes, ou en supras-segments qui subissent ce-ci, les proches d'un même type de son n'ont pas, dans cette langue.

Pour illustrer notre idée, à défaut de pouvoir la prouver, nous allons considérer trois faits. Le premier est le suivant: le système linguistique d'un locuteur interprète constamment dans la saisie qu'il a d'une langue seconde, en plus de sa propre langue. Notre expérience avec des informateurs de diverses langues confirme ceci totalement. Un arabo-phonie perçoit et réalise différemment /a/ et /e/ en français (belle - belle) mais ne perçoit pas, en arabe, la différence entre [a] et [e] dans [a] [e] "chien" et [e] "sorte d'âme", [a] et [e] étant des allophones dans cette langue. D'autre part, l'alternance phonétique arabe entre [a] et [e] est systématiquement perçue par les locuteurs présents dans des situations dans leur système et dans le françois. Le système linguistique de chacun intervient donc directement dans la saisie de sa propre langue, bien sûr, mais aussi dans la saisie et l'interprétation d'une langue seconde. Lorsqu'un montagnais prononce /f/, un francophone entendra /f/, tandis que /f/ est /f/, tandis que /f/ est /f/. Ce raccrochage du /f/ montagnais à trois phonèmes différents consiste, de la part du locuteur francophone, une pratique stricte de ses propres habitudes de discrimination. On pourrait multiplier les exemples. Même les variations sous-distinctives ne sont pas perçues et interprétées lorsque c'est utilie: [r] plutôt que [k], par exemple, sera considéré comme "archaïque" et "rural". Deuxième fait à considérer: l'attitude d'un locuteur à son langage est homogène et à l'hétérogène, au convergent et au divergent qui s'y manifeste, face aux variétés à l'intérieur de sa langue et face aux autres langues. Une façon intéressante d'aborder le problème est d'examiner la transcription phonétique telle que d'ont descripteurs différents opérant sur un même oups. L'exemple que l'on veut citer s'est déroulé dans le cadre d'un projet de recherche sur la langue en- francs francophones de France et du Québec. Deux Québécois et un Français ont transcrit une même partie de corpus oral. L'examen des trois transcription fait apparaître que le premier transcris- teur québécois a gommé tous les traits du phonétisme québécois (assibilation, diphtongaison, syncope,
dénasalisation, etc.) en modélisant sa notation sur le français standard. Le second Québécois a bien marqué les traits phoniques du québécois et a interprété le français en fonction de celui-ci (il surphonologise, par exemple: "un" /arp/. Le Français, quant à lui, a pris sa variété comme norme et a interprété le québécois comme déviant. Il a dégagé des traits québécois partout, même là où il n'y avait pas et a manqué de percevoir des distinctions qu'il ne possédait pas (/ɑ/ ~ /ɔ/). Le filtre personnel dépèr d'archétypes divers, conscients et inconscients, d'ordre culturel, fantasmatique et autre, module la saisie et la représentation de la langue. Il intervient dans la langue. Il est partie intégrante de la langue. L'attitude du locuteur face à la langue doit donc langue. Enfin, au niveau du modèle construit par l'analyste, le processus est toujours identique, peu importe la théorie sous-jacente. Les très nombreuses définitions des traits, du son, de la syllabe, du phonème, du prosodème, etc., attester parfaitement du fait que les unités dégagées résultent de découpages de réalités perçues et interprétées en fonction d'une pertinence. Les relevés faits sur les organes phonateurs ou le signal acoustique sont des abstractions faites à partir de point de vue, d'une perspective particulière. Les traits phonétiques sont des traits conçus, recensés à la réalité perçue, d'une représentation formalisée, le cas échéant. De même, les traits phonologiques sont fonction de la façon dont on perçoit les oppositions mais aussi de la théorie utilisée (concepts, procédures, méthodes) pour construire le modèle. Observation, description et explication ponctuent l'élaboration du modèle, cette image que tout veulent cohérente et la plus conforme possible à la réalité perçue. Certains besoins collectifs s'actualisent en langues par l'acquisition d'habitudes de discrimination incarnées chez les sujets. Les filtres perceptuels sont les nœuds de rencontre du social et de l'individuel où naissent les langues. Il n'y a pas de langue sans filtre et il n'y a pas de modèle sans langue. Langue, filtre et modèle sont indissociables, quoi qu'il en soit il existe beaucoup de langues non modélisées. Dès lors, s'il n'y a pas de différence de nature entre les traits phonétiques et les traits phonologiques, doit-on rejeter pour autant l'opposition entre phonétique et phonologique? Nous pensons que non.

Dans le modèle, les traits phonétiques et phonologiques jouent le même rôle. Ils constituent des représentations formelles de réalités perçues. Leur fonction est ici de type descriptif et explicatif, invalidant l'équation faite entre les deux couples phonétique/phonologie et substance/forme. Toutefois, au niveau des habitudes de discrimination acquises, manifestées et perçues, leur structure et leur fonction sont différentes. Les traits phonétiques sont des qualités articulatoires, acoustiques, ou auditives, positives, abstraction faite de la pertinence de chaque trait dans l'établissement de la communication pour une langue donnée. Les traits phonologiques sont, au contraire, des qualités négatives en ce qu'ils font intervenir avant tout le caractére opposite de chaque trait pour une langue donnée, établissant un ordre hiérarchie fonctionnelle des rôles exercés par les éléments phonétiques dans chaque cas. Au niveau de la langue, la dichotomie phonétique/phonologie apparaît donc comme fatalement invérifiable, par définition. Cependant, il faut insister sur la touche nécessaire et fatalément inévitable complémentarité des deux approches au niveau du modèle. Les unités linguistiques étant des opérations abstraites de base cognitive, il n'est pas insensé de considérer, à la limite, que toute langue est elle-même un modèle. On ne pourra jamais confondre pour autant les unités du modèle et les unités de la langue. Celles-ci ont une fonction explicative, celles-là ont pour fonction la communication.

Le filtre perceptif individuel est au cœur de la langue. Il en est indissociable. La langue englobe, l'intègre tout en s'actualisant par lui. Voilà le sens des pointillés dans la figure qui suit. La membrane qui sépare la langue du filtre est totalement poreuse, perméable. Après quoi, la langue pourrait être observée, décrite, analysée et à l'aide d'une théorie, représentée dans un modèle. Le caractère a posteriori et facultatif de la modélisation par l'entremise d'une visualisation formelle explicite est symbolisé ici par la ligne en traits. Les unités linguistiques du modèle restent des approximations rationnelles des unités linguistiques de la langue pour lesquelles seules les oppositions demeurent garantes.
PHONOLOGICAL UNITS IN SPEECH: UNRECOGNIZABLE ABSTRACTIONS FOR THE COMPUTER

Laurent Santerre
Linguistics Department, University of Montreal

The difference between orthographic and phonemic representation in language is clearly understood. The distance between phonemic and phonetic representation may be just a question of how much is well understood. People do not speak as phonology would have them speak. Intravocal and especially interfamidividual variation results in the production of a large number of allophones.

Three elements which must be carefully distinguished in order to describe speech precisely:

1) speech production by the speaker;
2) the acoustic (physical) effect of the speech signal;
3) psychoacoustic perception, or the listener's semantic interpretation.

Our knowledge of real speech signals is still largely symbolic, influenced by unwritten abstraction phonology and by perception. The ear is an effective decoder of meaning, but a rather primitive tool for acoustic analysis. Functional analysis of the speech signal is limited in that it does not enable us to identify real and pertinent speech parameters. When these analytic parameters are resynthesized, the analyzed speech is not faithfully enough reproduced.

The best form of analysis is synthesis. This may be partial or complete, or it may be unknown. High-fidelity synthesis demonstrates exactly how accurate the analysis is. Analysis-synthesis requires the constant involvement of the operator and though it is inconvenient, it is necessary if we are to understand real-time, the true functional parameters of speech. Sentences are now being perfectly synthesized at our laboratory and at the LIMSI laboratory in Orsay, France.

Here are a few examples that demonstrate why phonological and phonetic definitions of speech units are largely problematical and inadequate both for linguistic theory and automatic recognition:

1) disappearance of high vowels;
2) sonority of half-vowels;
3) allophones of /R/.

1) The Disappearance of High Vowels (Text)

When the high vowels /i, y, u/ are unstressed, they become considerably shortened in several languages. This is very common in Quebec French though it goes completely unperceived, even to language observers. I asked a student to read several sentences in a soundproof booth and then played them back to 35 phonetics students. The reader was not aware of what I was investigating, but the listeners were told the unstressed high vowels frequently disappeared.

The text was before them and they were asked whether they thought the underlined high vowels had been spoken or pronounced. They could hear the tape as often as they wished.

Here are the sentences, and the percentage of cases in which listeners thought they heard the high vowels. It should be noted that acoustic analysis (Kay sonagrams and much more precise computer-digitized sonagrams) did not register either the vocal or formant resonance of any of these vowels.

1) Il est populaire (64.33%). 2) Je vous dirons (61.48%) un support (74.32%) en bois. 3) Ça peut pas plaire à tout (100%) le monde. 4) Le sport de compétition (54.33%) est populaire (65.62%) de nos jours.
5) Pierre, je suppose (34.38%) que tu (85.73%) vas nous (94.32%) accorder ton support (75.73%) financier (100%).
6) Il s'installe la biais de discussion (70.65%) (85.37%). 7) Tu as encore oublié tes billets, je suppose (37.17%).

Clearly, the word the word is understood in context, we think we hear it in its entirety because we know its spelling or phonology. Even when the word is degraded by production and perception must therefore be clearly differentiated, and the acoustic signal must be studied in itself. I demonstrated this indirectly by playing the same words out of context and in a random order: populaire, distingué from pas plaire, nor support from sport.

2) The Sonority of Fricatives

I asked several students, some of whom were European Francophones, to pronounce words such as girouette and gazon. In 92% of cases the fricatives were not voiced but were perfectly intelligible. In girouette for example, even when listeners tried to perceive the frication as a /j/, they were unable to do so. It had been heard as a rearticulated frication as a /j/. When the Québécois word chirade (derived from the English sheath) was heard in segments, its /j/ simply seemed more strident in comparison. The feature that distinguishes /j/ from /j/ is not sonority. In some cases it may be the increased duration of the /j/ and when duration is comparable, slight turbulence around 500 Hz for the /j/ because of a partial meeting (without vibration) of the vocal chords (Fig. 1). This lack of vibration may be caused by powerful airflow in the glottis or intense supra-glottal pressure created by great constriction at the point of articulation.

We transcribed words pronounced by a Broca's aphasic suffering from phonemic disintegration. These words all had the phonology refers to as a voiced fricative at the absolute final position. Acoustic analysis, however, showed that none of these consonants was even partially voiced. The test results are as follows.

The word /z/ was recognized in 93.7% of cases (and interpreted as /z/ in 6% of cases), because the vowel was sustained and not lax as is the norm in Quebec French when a high vowel is closed by a shortening consonant. Half of the listeners observed that the /z/ was not voiced, whereas because they were specifically asked to judge the sonority of the final fricatives. The word /c/ was recognized in 100% of cases: without a minimal pair listeners had no choice, and 46.6% found the fricative more or less unvoiced. The word /c/ was intelligible in only 8.4% of cases, and was quite predictably heard as /ʃ/. In the Quebec phonological system, the /o/ is long by nature and cannot be shortened by coarticulation of the shortening consonant that closes it. There was no reason in this instance not to hear the final /ʃ/.

Conclusion: We should not trust perception to define the distinctive feature of speech.

3) /R/ Allophones (Fig. 2 to 9)

Il faut se taire. (Fig. 2). This /R/ is formed by retracting the tongue in the pharynx without constriction: it is thus vocalic and not consonantic in nature. It is immediately followed by an apical flap. This double allophone is common in Montreal (Radio-Canada) and seems to be a stage in the replacement of the apical /R/ by the posterior /R/ (Santerre 1980 and 1982). The following observations are based on the study of vocal tract X-Rays of this working-class speaker (Fig. 2 and 3).
Regarde la reine (Fig. 3). This short sentence demonstrates three types of /r/ usage: the multiple-flap apical /R/; the posterior vocalic /R/ followed by a single apical flap; and the intervocalic /R/ with one apical flap. The intervocalic /R/ is always a single apical flap in this speaker's system.

Rosaire (Fig. 4). This initial /R/ is a voiceless aspiration. The speaker is a boy from Havre-St-Pierre on the northeastern shore of the St. Lawrence.

Une sorcière (Fig. 5), same speaker. Children from the same school in Havre-St-Pierre form consonant-preceding /R/ within words as an occlusive; this occlusion may be formed in the front, center, or back and velar part of the palate.

Ils sont gris (Fig. 6). The single-flap apical /R/ is necessarily separated from the preceding consonant by a vocalic element. The speaker is a nine-year-old boy from a working-class Montreal neighbourhood where the /R/ is usually anterior.

De la truite (Fig. 7), uttered by a girl from Havre-St-Pierre. This consonant-following /R/ is a bilabial fricative and unvoiced velar.

Des crables (Fig. 8), by a boy from Havre-St-Pierre. Here the /R/ is a velar fricative difficult to distinguish from the preceding lax /k/.

Une carabine (Fig. 9), same speaker; the intervocalic /R/ is imperceptible in this word.

More allophones cannot be presented here. Many are found in addition to the usual anterior and posterior /R/ as, such as the retroflex English /R/ (soeur, sportive), the word-final /r/ a dipthongised in the anterior series (faire [faj]), the vocalic /R/ formed solely by the convergence of forms 1 and 2 without constriction, and the double /R/, which is both posterior and anterior.

Conclusion: Phenomenal units are abstract units, not formed according to the features and indices that define them. Automatic recognition is only possible when the many conditioned allophones of each phoneme are in the computer's memory, and this is only for a single speaker. Phonemes are defined to exclude each other reciprocally, but their allophonic zones overlap considerably even for phonemes theoretically defined as very distant. To be recognizable, language sounds must be taken for what they are and not for what we believe they should be when we wish to render them as recognizable as possible.

References
The acquisition of Cantonese phonology: a phonetic analysis

S.-M Tse1 and D. Ingram2

1Silverman Hearing Research Laboratory, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, M5G 1X5 Canada
2Department of Linguistics, University of British Columbia, Vancouver, B.C., Canada

As part of a general study on the acquisition of Cantonese (Tse, 1982), this paper presents a phonetic analysis of three children learning Cantonese as their native language. "Wai", the primary subject in the study, was observed longitudinally for a period of one year. "Wing" and "Ching", the other two subjects, were observed cross-sectionally for the purpose of comparison. All three children were born and brought up in exclusively Cantonese-speaking families in Canada.

We begin with a brief description of the Cantonese phonology. In Cantonese, the syllable structure is simple. It is traditionally described by Chinese phonologists as having three parts: (1) the initial, or first consonant, (2) the final, and (3) the tone (e.g., Chao, 1947). In Cantonese, it is possible for a syllable to have no initial segment, and is then said to contain the 'zero' initial. The final consists of a vowel, or a vowel followed by a consonant or a glide. It may also be a syllabic nasal /ŋ/ or /ŋ/ if it is the only constituent in the syllable.

Central to our present study are both the initial consonants and finals. They are listed in Tables I and II respectively. The description is mainly based on Wong (1940) and the observation of the speech of the subjects' parents.

Table 1: Initial Consonants of Cantonese

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Unaspiratedstops</th>
<th>Aspiratedstops</th>
<th>Nasals</th>
<th>Fricatives</th>
<th>Sibilants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labials</td>
<td>p</td>
<td>p'</td>
<td>m</td>
<td>f</td>
<td></td>
</tr>
<tr>
<td>Dentals</td>
<td>t</td>
<td>t'</td>
<td>n</td>
<td>l</td>
<td></td>
</tr>
<tr>
<td>Alveolars</td>
<td>ts</td>
<td>ts'</td>
<td>s</td>
<td>j</td>
<td></td>
</tr>
<tr>
<td>Velars</td>
<td>k</td>
<td>k'</td>
<td>h</td>
<td>w</td>
<td></td>
</tr>
<tr>
<td>Labialized-velars</td>
<td>kw</td>
<td>kw'</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The parents of our three subjects were all immigrants from Hong Kong. At the time of the study, they had been settled in Canada from 4 to 7 years. Although all subjects were born in Canada, they came from homes in which only Cantonese was spoken. Wai, the primary subject, was visited biweekly from the ages of 1;7(14) (one year seven months and 14 days) to 2;8(11). This resulted in a total of 25 sessions. There were a total of 7292 recorded utterances. The speech samples of Wai were collected at ages 0;1(10), 1;6(2), 1;8(5) and 2;2(29), and Wing at ages 0;4(20) and 2;0(96).

The tapes were transcribed by the first author (Transcriber 1) and a university student. Both are native speakers of Cantonese, and have training in phonetic transcription. All the tapings of the odd-numbered sessions were first transcribed by Transcriber 1 and were double-checked by Transcriber 2. The transcription order a sound must be 'acquired' or 'transitional' in the three other sessions. Articulation Scores are shown in parenthesis. As indicated in Table I, one major problem that an effective phonetic analysis needs to solve is the question of frequency. For example, if Wai produces 3 cases of syllable initial [p] and 20 of [p'], should we say that she produces both [p] and [p'], should we ignore [p] because of lack of frequency, or should we include [p] but give [p'] special status? Following Ingram (1981), this study adopted a criterion of phonetic frequency in Cantonese which classifies the sounds of the language into 3 categories: (1) frequent, (2) used, and (3) transitional. This criterion is based on the total number of consonants and vowels in the language. Two measures of phonetic ability are calculated. The first, the Total Number of Sounds in the language, includes the sounds that meet the criterion of frequency. The second measure is the Articulation Score which is a more precise measure than the first one. This measure assigns points for each frequent sound, 2 points to those which are used, and 1 point to transitional sounds (see Tse, 1982 for a detailed discussion).

Table III lists the division of Wai's sound segments into those 'acquired' and those 'not acquired'. To be 'acquired', a sound must be 'acquired' or 'transitional' in the three other sessions. Articulation Scores are shown in parenthesis. As indicated in Table III, one sees that for initials consonants, except for the unaspirated labialized velar stop /kw/, all the unaspirated stops and affricates were acquired. On the other hand, none of the aspirated stops and affricates were acquired. The reason that the unaspirated labialized velar stop /kw/ is not acquired is probably due to the infrequent occurrence of this sound in the language system. /kw/, together with the labialized velar stop /kw/, can only appear together with a limited number of finals (Hashimoto, 1972). Other sounds that were not acquired are initial /ŋ/, /j/, /w/ and /j/.
As for the final consonants, two out of three nasals were acquired, i.e., /n/ and /ŋ/. Also there is only one final stop that was acquired, i.e., /k/. It seems that Waï has a slight nasal preference. As for the vocalics and syllabic consonants, one sees the basic vowels were acquired, except for the two front rounded vowels /y/ and /ø/. However, in terms of diphthongs, only 2 out of 10 diphthongs (20%) were acquired. They are /ai/ and /ui/. Syllabic /ŋ/ was acquired but /ŋ/ was not.

Table III: Division of Waï’s segments into those “acquired” and those “not acquired”

<table>
<thead>
<tr>
<th>Initial consonants</th>
<th>Final consonants</th>
<th>Vocalics &amp; Syllabic consonants</th>
</tr>
</thead>
<tbody>
<tr>
<td>m(10)</td>
<td>n(12)–ŋ(14)</td>
<td>s(17) w(12)</td>
</tr>
<tr>
<td>p(11)–t(14)</td>
<td>k(16)</td>
<td>a(17)</td>
</tr>
<tr>
<td>ts(15)</td>
<td>s(13)–h(12)</td>
<td>y(9) w(11)</td>
</tr>
<tr>
<td>j(10)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Based on the Articulation Scores for the sound segments in Table III, a hierarchy of the frequency of occurrence of individual segments for Waï was set up. It is presented in Table IV. Its purpose is to see which sound segments are more preferred, and which ones are less preferred or not attempted by Waï.

Table IV: A hierarchy of the frequency of occurrence of sounds for Waï in terms of Articulation Scores

<table>
<thead>
<tr>
<th>Rank</th>
<th>Initial consonants</th>
<th>Final consonants</th>
<th>Vocalics &amp; Syllabic consonants</th>
<th>A.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>m</td>
<td>n, ŋ</td>
<td>s, w</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>p</td>
<td>k</td>
<td>a</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>ts</td>
<td>s</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>t</td>
<td>ŋ</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>s</td>
<td>h</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>6</td>
<td>h</td>
<td>n</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>p</td>
<td>k</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>8</td>
<td>j</td>
<td>s</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>9</td>
<td>m</td>
<td>ŋ</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>n</td>
<td>s</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>11</td>
<td>ts</td>
<td>h</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>p</td>
<td>j</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>k</td>
<td>s</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>14</td>
<td>kw</td>
<td>ŋ</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>15</td>
<td>s</td>
<td>m</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>f</td>
<td>ŋ</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>ŋ</td>
<td>s</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

From Table IV, one sees that among vocalics and syllabic consonants, /i/, /a/, /e/ and /ɛ/ are the most frequent, whereas /ai/, /ui/, /y/ are the least frequent. Among initial consonants, /ts/, /t/, /k/ and /ŋ/ are the most frequent, whereas /m/, /ŋ/ /ŋ/ /k/ and /ŋ/ are the least frequent. Among final consonants, /ŋ/, /n/ and /k/ are the most frequent, whereas /s/ /h/ /t/ are rather low, ranging from 2 to 0. This indicates that Waï mainly used /ŋ/, /n/, and /k/ finals in her speech, while the other three finals were seldom used.

From the data presented above, it is possible to set up an inventory of the first sounds acquired in Cantoneese. We conclude with an inventory of the early sounds ‘used’ by Waï, Ching, and Wing. To do this, we have taken sounds that appear as ‘used’ in the phonetic inventories of at least two subjects. These are shown in Table V. Those sounds ‘used’ by all three subjects are circled to separate their more secure status in the analysis. Data like that in Table V can be used for comparison to that from other Cantoneese subjects to see the basic sounds acquired, and also to data from children learning other languages.

Table V: Summary of sounds acquired by Waï, Ching, and Wing.

<table>
<thead>
<tr>
<th>Initial Consonants</th>
<th>Final Consonants</th>
<th>Vocalics &amp; Syllabic consonants</th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td>n, ŋ</td>
<td>s, w</td>
</tr>
<tr>
<td>ts</td>
<td>h</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

REFERENCES

Wong, S. L. 1940. Yue-yin yun-hui (A Chinese syllabary pronounced according to the dialect of Canton.) Hong Kong: Zhong Hua Shuju.
TEMPORAL FACTORS IN VOWEL PERCEPTION

Anton J. Rozylyal

Department of Linguistics, The University of Alberta
Edmonton, Alberta, Canada T6G 2E7

In the frequency range of the first formant all harmonic components of the periodic vowel spectra are separated by frequency intervals wider than the critical bandwidth and thus can be resolved by human listeners.

Several vowel perception studies addressed the problem of how listeners estimate the frequencies of the first formant, that is, the positions of spectral envelope peaks. From the harmonic components of the vowel line spectra. Of particular interest is the case when the actual resonance frequency of the vocal tract is not an integer multiple of the fundamental frequency. In such case of the spectral lines coincides with the formant frequency, which is commonly referred to as the "missing formant energy" problem.

Mushnikov and Chistovich (1971, 1973) proposed that only the most prominent harmonic component from the first formant range determines the perceived formant frequency. Carlson, Pant, and Granström (1971) concluded that the listeners perform a weighted average of the two most prominent spectral components in order to estimate the frequency spectra. The task of the formant frequency detector is presented in the spectral domain, in terms of the line frequency spectra of the stimulus. This indirectly assumes not only that the signal is sufficiently long for the individual spectral lines to develop but also that the analyzing window of the auditory analyzer is equally long for the spectral lines to be extracted. All these models have in common the implicit assumption that the vowel recognizing mechanism first extracts from the vowel signal the amplitudes and frequencies of its harmonic components in a given frequency range. Out of these components this mechanism then selects a number of prominent components, and then, using some weighting process of these prominent components, estimates the formant frequency on which the vowel recognition depends. This multistage process, apart from its temporal aspects to be discussed shortly, is wasteful since it involves extraction of more information from the signal than is needed.

Let us now consider temporal aspects of vowel recognition, that is, let us try to estimate the duration of the analyzing window of the processes which may be involved. One such process is vowel recognition, another in detection of periodicity in the vowel signal, another still, estimation of the frequency of individual spectral lines.

As can be estimated from the data available in the literature, the time interval needed for vowel recognition is of the order of one glottal period (Gray, 1942). Results of several earlier studies support this claim.

For instance, Hyde (1969), using natural vowels produced by a male speaker, found that for most of the front vowels one glottal period of the vowel signal is sufficient for correct recognition. Most of the remaining vowels were correctly recognized when about two periods long. The remaining misidentifications could be expected due to the removal of the duration information.

Powell and Tosi (1970) found the mean temporal threshold of recognition for various vowels to be between 9.3 ms and 27 ms, with a mean value of 15 ms. At the voicing frequency of 125 Hz they used, this corresponds to a mean vowel duration of just below two glottal periods.

Moore and Mundie (1971) measured the minimal number of glottal periods required for reliable vowel identification. Their vowel segments, extracted from sustained vowels produced by a male speaker, were between 4 and 40 ms long, which corresponded to between 1/2 and 5 glottal periods. They obtained no substantial improvement of the recognition rate if the vowel duration was extended beyond two periods.

Suon and Beddoo (1972) measured the effect of learning on identification of six synthetic vowels or 10 ms duration. At the male voice fundamental frequency used, all stimuli contained about 1.3 vowel periods. Two groups of subjects with different amount of phonetic training participated in the study. For subjects with no phonetic training all vowels sounded initially like clicks. Most of the experienced subjects recognized the low vowels immediately. Over the learning period the average correct recognition rate of the phonetically experienced subjects increased from the initial 51 to 92%. Phonetically naive subjects improved their recognition from 2% to 74%.

All these short vowel perception studies agree in finding that the recognition in general is poorest for the high vowels with the lower first formant frequencies. This can be explained by assuming that for these vowels the number of formant periods which would fit into the short analyzing window is too small for sufficiently accurate frequency estimation.

More recently, Nearney and Assmann (1981) investigated the role of inherent spectral changes in the perception of isolated Canadian English vowels, including vowels generally regarded as monophthongal. Their stimuli were constructed by excising two 30 ms segments from the vowel signal, one from the vowel nucleus, the other from the vowel offglide. Both segments were shaped by a Hamming window so that for a male voice each segment in effect contained about two glottal periods. These segments were presented to the subjects for recognition in either natural or reverse temporal order, with a 10 ms silent interval inserted between the two segments. The original unmodified vowels were also tested. The experimental results demonstrated that the offglide portion contributes significantly to recognition. While the recognition rates for the unmodified vowels and the two-segment vowels in natural order were comparable, amounting to 89% and 84% respectively, the two-segment vowels in reverse order was only 63%. These results indicate that the subjects detected both the identity and temporal order of the two vowel segments. This suggests that in analyzing the vowel signal the recognition mechanism does not include both segments in one analyzing window, and that means that this window must be shorter than about 40 ms.

The just mentioned data in summary indicate that a vowel signal of a duration between one and two
glottal periods is sufficient to yield recognition rates which are not much lower than those found for sustained vowels.

A signal only two periods long cannot be recognized as periodic. A listener needs several periods of a periodic signal to detect its periodicity and to establish its pitch. Doughty and Garner (1947, 48) investigated pitch characteristics of short pure tones. From their data it can be estimated that a pure tone from the glottal frequency range must be about 40 ms long to be perceived as a so-called "click-pitch", that is, to reach the duration threshold for tonality at which it acquires a tonal quality accompanying the perception of a click. That means that for a male voice range the stimulus must be at least five periods long to have a tonal quality. For a female voice range that duration threshold of tonality is about seven periods.

It takes even longer than that to bring the process of the extraction of individual spectral lines to completion. Korn (1962) showed that a pure tone of arbitrary frequency with an optimally shaped envelope must be at least about 50 periods long to produce a click-free pure pitch sensation. Experiments on frequency discrimination as a function of stimulus duration conducted by Bahls (1948), Oetinger (1959), Liang-Chian and Chistovich (1961), Carluccio (1962), Rodden (1971), and Moore (1973) indicate that frequency discrimination of pure tones in the glottal frequency range improves for about 200 ms. This interval is compatible with the duration of a typical vocal nucleus. Moore (1973) designed his experiment to find the frequency range in which the competing "place" and "temporal" mechanisms are more effective in extracting the frequency information from short pure tone stimuli. He found that in the low frequency range the frequency discrimination is by one order of magnitude better than the discrimination which can be provided by the "place" mechanism. Moore concludes that all the evidence indicates that a "temporal" mechanism operates for frequencies below 5 kHz. This range includes the first two vowel formants. Any efficient time-measuring mechanism must have a good temporal resolution and thus use a short temporal window.

With this experimental evidence indicates that vowels can be recognized sooner than the listener is able to tell whether the signal is periodic or not. As can be estimated from the experimental data listed above, vowel recognition requires about 40 ms (one period of a female voice) and 16 ms (two periods of a male voice). On the other hand, detection of periodicity in a signal of voice frequency requires about 40 ms. To extract full frequency information from the individual spectral lines takes even longer, about 200 ms.

Spectra of vowels analyzed with a window one glottal period long are continuous, with peaks exactly at formant frequencies. These short-term windows do not provide a significant amount of information. A pass filter with a time window shorter than one glottal period cannot resolve individual spectral lines due to the constraints imposed by the uncertainty principle (Gabor 1946, 1947). If the frequency resolution of a filter is better than the frequency separation between two adjacent formant frequencies, only formant frequencies appear in its response. These are amplitude modulated, with a sharp rise and gradual exponential decay, following the periodicity of the glottal source. The mechanical filter of the basilar membrane with its frequency resolution approximately one third octave wide appears to be well suited for this type of analysis.

All this suggests that spectral energy at the formant frequency is directly available to the listener. Another point which can result from this discussion is that more attention in vowel perception should be directed towards the time domain, both in describing the vowel stimuli and in modeling the recognition mechanism. If synthetic vowels are used as stimuli, care should be taken to simulate properly the exponentially decaying response of individual formants. This is always the case when using terminal type synthesizers. But when generating the formant responses from individual spectral lines, it is necessary to preserve proper phase relationship between them.

Neurophysiological evidence also supports this time domain approach. Young and Sachs (1979) investigated the representation of steady state synthetic vowels in the temporal aspects of the response patterns of single auditory nerve fibers. Interval histograms of these discharges show that for stimulus sound pressure levels of about 80 dB the responses occur almost exclusively at the first and second formant frequencies, at their harmonics and at their intermodulation products. This indicates that the formant frequencies are extracted at the peripheral stages of the auditory nerve.

It can be concluded that in vowel recognition task the temporal characteristics of events is as follows: the vowel quality is recognized by an analyzer with a window one or two glottal periods long. This analysis may be repeated over the whole duration of the vowel stimulus and the individual readings averaged in order to improve the accuracy of formant determination in a similar manner to that proposed by Liang-Chian and Chistovich (1961) for the case of frequency discrimination. For this type of analysis frequency resolution as provided by the basilar membrane appears to be sufficient. If demanded by the testing paradigm, the subject can estimate the fundamental frequency of the speaker's voice or resolve all of the harmonic components of the long-term vowel spectrum in the first formant range, but this task appears to require longer time interval, about the whole duration of the vocal nucleus.

ANALOGS TO SPEECH PROCESSING: 
DISCRIMINATION OF DYNAMIC, COMPLEX SOUNDS 

E.B. Slawinska, D.G. Jamieson, and T. Johnson 

Speech and Audition Laboratory, Department of Psychology, The University of Calgary, 2500 University Dr. N.W., Calgary, Alberta, Canada 

Several phonetic distinctions seem to depend on specific auditory constraints. One type of auditory constraint is related to the rapid changes in formant frequency (formant transitions) which occur in voiced consonant sounds when the vocal tract changes shape while the vocal folds are active. 

Two aspects of formant transitions are especially important: their direction and their duration. Transition duration cues the manner of speech production (for example, the /ba/-/wa/ stop/semi-vowel distinction). The direction of formant transitions cues consonant place of articulation, distinguishing /da/, /gi/ and /ba/, for example. These transitions all take approximately 50 ms, but /da/ is cued by an initial, falling second formant (F2) transition, /ga/ by a more steeply-falling F2, and /ba/ by a rising F2. 

Experiments with sine-wave analogs of formant transitions have studied performance when the initial and final frequencies of the transition are fixed, while transition duration is varied. Discriminability follows Weber’s Law for transitions alone, with the fastest transitions discriminated best (Jamieson and Slawinska, 1984). However, when transitions are followed by a pure-tone (analogous to the steady state formant of the vowel part of the CV syllable), d’ values are highest for glides in the 40-60 ms range (cf., Jamieson and Slawinska, 1984). This effect depends on the intensity/time parameters of the steady state, however. The point of maximum discriminability shifts to faster transitions as the amplitude of the steady state is reduced, suggesting that the steady state normally masks the transition. 

The present experiments a) examined the discriminability of transition/steady state stimulus complexes, when the signals were periodically, rather than continuously excited, and b) examined the possible role played by the steady-state vowel in masking transition cues in speech. 

Method 

Our isolated formant stimuli varied in duration from 10 msec to 90 msec, in 10 msec steps, while the duration of the steady-state covaried from 190 to 110 msec to maintain an overall stimulus duration of 200 msec. The “voiced” set of stimuli was synthesized to follow the center frequency of the isolated F1 track of a /ba/ sound, with voicing throughout the signal and bandwidth of 90 Hz. The “sinewave” set of stimuli followed the same center frequency tracks, but consisted of a continuous tone glide followed by a pure tone. Stimuli were presented at a comfortable listening level (68 to 83 dB), adjusted separately for each subject, at the beginning of the experiment. 

For the second experiment, we synthesized a seven-item stimulus series which ranged from a clear /bad/ to a clear /wad/, by varying the duration of the initial transitions of the first two formants in 5 ms steps between endpoint values of 15 ms and 45 ms. From this original series we created four additional series by attenuating the steady-state “vowel” portion of each member of the continuum either totally or by 10 dB, 20 dB, or 30 dB relative to the original amplitude. The attenuated interval began 10 ms into the steady-state segment and ended 5 ms before the beginning of the final formant transitions. Stimuli with attenuated steady-states were audibly different from the originals, but the phonetic identity of the components (including the vowel) was preserved (cf., Strange, Jenkins, and Johnson, 1984). 

In the first experiment, our listeners were required to indicate whether the pairs of sounds were the same or different. 

For the speech experiment, the stimuli from all five series were randomly ordered and tape-recorded to form an identification test, which was played binaurally via earphones to a group of 20 undergraduates with normal hearing. Their task was to identify the initial consonant on each trial as /b/ or /w/. 

Results and Discussion 

Performance in the transition discrimination task was summarized by computing the discriminability statistic -ln n (Luce, 1963), which is monotonically related to d’. The figure shows representative functions relating discriminability to transition duration. Sinewave stimuli with glide durations in the range 40-60 msec were best discriminated. For each listener, voiced stimuli were less well discriminated, and they showed a single discrimination peak at a faster transition rate.
Discriminability of differences in transition duration for sinewave (dashed lines) and voiced (solid lines) for individual listeners. Note the general tendencies for a significant peak to occur in the 40-60 ms range for sinewaves and at a somewhat faster transition rate for voiced stimuli.

The results of the speech identification task showed that the location of the phonetic boundary between /b/ and /w/ depended on the degree of attenuation of the steady state. In the original series, the group phonetic boundary was located at a transition duration of about 32 ms, a value which agrees well with those obtained previously using pure tone glides or single formant stimuli. In the Total Attenuation series, however, the /b/-/w/ identification boundary shifted to a transition duration of about 26 ms. For the three series having less than total attenuation of the steady state segment, shifts in the identification boundary were smaller and less reliable. These results indicate that fundamental psychoacoustic processes like backward masking play a role in identification of phonetic segments distinguished by temporal cues such as the duration of formant transitions. For example, backward masking may help explain why /b/-/w/ identification boundaries shift as a function of the duration of the steady state segment which follows initial formant transitions. Postulation of more complex processes such as rate normalization may not be necessary to account for such phenomena (cf., Miller and Liberman, 1979).

General Discussion

The demonstration that sensitivity to frequency transitions is enhanced when the transitions occur over an interval of approximately 40 ms helps to explain why speech formant transitions occur over a similar temporal interval. This seems to be a reliable effect for stimuli with periodic excitation, such as that found with voiced speech sounds, as well as for the continuous tone glides which we used in our earlier experiments. It seems reasonable to conclude that the auditory system is specially prepared to respond to transitions within this interval. These results may help to explain how human infants are able to use the information contained in formant frequency transitions to discriminate speech sounds. One explanation for these and other related results is that in the process of the evolution of language, the sounds of the speech repertoire were selected to be maximally discriminable to the human ear. Evidently, a portion of this selective pressure reflects intra-speech masking effects.

*This research was supported by grants from the AID/FR, HFC, and NSERC to DGJ, and by Fellowships from the AID/FR awarded to EBS. We are grateful to R. Assmann, M. Chassman, V. DiLillo, T. Nearcy, and A. Rozsas for their comments on this work.

References


ON THE ROLE OF SPECTRAL TRANSITION IN PHONEME PERCEPTION AND ITS MODELING

S. Furui and M. Akagi

NTT Electrical Communication Laboratories
Musashino-shi, Tokyo, 180 Japan

INTRODUCTION

Continuous speech can be characterized by time-varying spectral patterns which have almost no steady-state period. The spectral transition is regarded as playing an important role in speech perception. It can be considered that spectral undershoot, which is produced by incomplete articulation as a consequence of the coarticulation, is compensated for in the auditory system based on the spectral transition, and, therefore, that the spectral target value is predicted by the perceptual mechanism.

However, the importance of dynamic and static features in speech production has not been clarified, nor has the auditory mechanism for phoneme perception based on the dynamic features been clarified. This paper investigates the acoustic characteristics for the dynamic features through a comparison of mathematical modeling and hearing experiments.

ROLE OF SPECTRAL TRANSITION IN SYLLABLE PERCEPTION

Japanese CV (consonant-vowel) syllables uttered by two trained female speakers and truncated initially or finally at various positions were presented in random order to four female listeners for identification tests. For each syllable, the position where the syllable identification score as a function of the truncation position exceeded 80 % for the first time and where the identification score changes abruptly was detected and termed "the perceptual critical point" for the initial and final truncation.

In order to compare the results of identification tests with physical features, a spectral transition measure was defined for representing the magnitude of spectral dynamics using the regression coefficients of the cepstrum time function. The 1st through the 10th order cepstrum coefficients are extracted every 5 ms, and the 30 ms-length time functions are approximated using linear regression functions. The transition measure at time t, $D(t)$, is then defined as

$$D(t) = \left( \sum_{i=1}^{10} a_i \right) / 10,$$

where $a_i$ is the regression coefficient for the i-th cepstrum.

Since the regression coefficient corresponds to the linear variation of the spectral envelope pattern in a unit time represented by the cepstrum, $D(t)$ corresponds to the variation of the same cepstrum. The interval of 50 ms for extracting the $a_i$'s seems adequate for preserving the transitional information associated with changes from one phoneme to another. $D(t)$ is calculated every 5 ms by shifting the 50 ms-length window.

The time function of the spectrum transition measure was compared with the identification score as a function of the truncation position. The results averaged over all syllables are presented in Fig. 1. They show that the perceptual critical points can be related to the maximum spectral transition positions. The percent correct identification for the truncated syllables changes rapidly depending on whether or not the truncated syllable preserves the maximum transition position [1]. This implies that a speech wave segment of approximately 10 ms in duration, which includes the maximum spectral transition, is considered to bear the most important phonetic (consonant and vowel) information.

Consonant and vowel identification scores for final truncation are highly correlated, which means that vowel and consonant information mutually interact in the above-mentioned 10 ms period. This suggests that consonants are mainly perceived by a syllable unit based on the spectral transition into the vowels which follow them.

Another identification experiment in which both the initial and final syllable segments were deleted indicates that a speech wave of approximately 50 ms in duration including the maximum spectral transition at its middle position contains sufficient information for phoneme perception (see Fig. 1).

VOWEL TARGET PREDICTION MODEL

The experimental results described in the previous section are applicable to the auditory system as the spectral transition target to the dynamic features of the 50 ms period. This section proposes an auditory model for this mechanism, as outlined in Fig. 2, which is based on the psychological and physiological knowledge of the auditory system [2].

This model introduces the functional difference of the outer and inner hair-cells reported by Dallos [3]. That is, these hair-cells are respectively the sensors for the vibration as well as the velocity of the basilar-membrane. Moreover, the former are connected to both the afferent and efferent fibers, whereas the latter are connected to only the efferent fibers.

Additionally, this model includes the following auditory characteristics:

(a) The dynamics of not only formants but the overall spectral pattern are perceived.

(b) Perceptual switching from the preceding to the following vowel (categorical perception) occurs at the maximum spectral transition position (this is based on the above-mentioned experiments).

Fig. 1 - Relationship between the truncation position and identification scores for the truncated syllables. Time axis is indicated relative to the critical point of final truncation. Ti, Tf: perceptual critical point for initial and final truncation respectively, Tm: maximum spectral transition position.

Fig. 2 - Auditory model for vowel target prediction mechanism.
(c) The amount and velocity of the spectral transition compensate each other to preserve the quality of perceived phonemes.

(d) The spectral transition can be represented by a 2nd-order critical damping model.

(e) A speech wave of approximately 50 ms in duration is used for the target prediction.

Considering these points, the following equations are assumed for the dynamics of spectrum y:

\[
\dot{x} = -k y + \dot{y}
\]

\[
x = \lambda (x + k b)
\]

where \( b \) is the target value, \( k \) is a feedback factor and \( \lambda \) is a reciprocal time constant for the transition.

When \( k = \lambda \), the 1st-order system equation (3) is represented by

\[
\dot{y} - 2\lambda y + \lambda^2 y = \lambda^2 b
\]

which is the 2nd-order critical damping model. The solution of this equation is

\[
y = a (1 - \lambda t) \exp(\lambda t) + b,
\]

that is,

\[
x = a \exp(\lambda t) - \lambda b
\]

by substituting Eq. (5) into Eq. (2).

Therefore, let us assume a certain value for \( \lambda \) and calculate the value for the spectral sequence of \( y_n \) such that

\[
x_n = -\lambda y_n + \dot{y}_n.
\]

Then,

\[
\epsilon(\lambda) = \sum_{n=N/2}^{N} (x_n - \dot{y}_n)^2
\]

can be minimized using the least mean square error method for

\[
\dot{\lambda} = -\lambda \exp(\lambda n) - \lambda b
\]

by varying the parameters \( \dot{\lambda} \) and \( \lambda b \). Next, \( \lambda \) can be optimized by minimizing \( \epsilon(\lambda) \) as a non-linear optimization problem. The optimum \( \lambda, a \), and \( b \) are obtained as the condition which minimizes \( \epsilon(\lambda) \).

This method solves the parameter values of \( \lambda, a \), and \( b \) for the 2nd-order critical damping model by solving the parameter estimation problem of the exponential function which corresponds to the 1st-order model. Since the initial transition position does not need to be decided by this model, the parameter prediction based on a short-period spectrum sequence was realized by the model, which has been impossible to do using the previous methods.

Although \( \dot{y} \) essentially represents the time derivative of spectrum \( y \), the 1st-order regression coefficient over the 50 ms-long period was used instead of it, taking the auditory time resolution and parameter stabilization into consideration.

EVALUATION OF THE VOWEL TARGET PREDICTION MODEL

Speech wave, time functions of the original spectrum \( y \) as well as the target \( b \) predicted by the proposed vowel target prediction model and the spectral transition measure are shown in Fig. 3 for the diphthong /a/. This figure clearly indicates that the target is reliably predicted and categorically changes at the maximum spectral transition position.

Figure 4 indicates sequences of the physically identified vowels whose ideal spectra feature the smaller spectral distances than the other vowels from spectrum \( y \) or from target \( b \) at each short period during the utterance of the Japanese word /kaa/. The center spectrum of each vowel uttered in isolation was used as the ideal spectrum for each vowel. Experimental results for 20 different words including every kind of vowel concatenation uttered by four male and female speakers show that the lengths of the transitional sounds, such as /e/ in Fig. 4 which is actually not perceived, become much shorter for the sequence of target \( b \) than for that of the original spectrum \( y \).

The perceptual critical point where the perceived vowel changes from /i/ to /a/, and, therefore, where vowel /a/ is initially perceived is also shown in Fig. 4. It was observed that the perceptual critical points correspond well to the points where the physical distance between target \( b \) and the ideal spectrum of the actual vowel attains a smaller value than any other vowel.

DISCUSSION

This paper has confirmed the importance of spectral dynamic features in speech perception, and has proposed an auditory model to predict the vowel target based on the dynamic characteristics. This auditory model appears readily applicable to phoneme segmentation and continuous speech recognition [4].

REFERENCES


PITCH OF GLISSANDOS IN SPEECH SOUNDS

C. Cavé, D.J. Hirst and M. Rossi
CNRS UA 261 "Parole et Langage", Institut de Phonétique, Université de Provence, 13621 AIX (France)

While a glissando by definition never has a fixed frequency, it is possible to perform pitch matching between a glissando and a pure tone. Studying the point of subjective equalisation or Pitch Point (PP) of glissandos (Rossi 1971) is a way to investigate the mechanism of the temporal integration of frequency.

A frequency change must present certain frequency and temporal characteristics in order for it to be perceived as a glissando. Frequency variations too small or too short are perceived as static. Nishihara (1979) using the results of Rossi (1971) showed that a satisfactory approximation of the glissando threshold can be obtained by a Gompertz function. The threshold curve divides the perceptual space below the curve are the infraliminal glissandos, above the curve the supraliminal ones.

Research on the value of the PP has shown that, for infraliminal glissando, this point is located between the 6/10 and the 7/10 of the glissando. These data support the hypothesis that the final part of a glissando could have a more important weight in the perceptual integration (Bradley & al 1941, Nabelek & al 1970, Tsunura 1976).

For the PP of larger frequency variations, however, subjects are able to perform equalisations with either the beginning of the glissando or with the end.

While the majority of previous research was based on pure tones, Rossi (1971, 1970) using speech sounds, found a PP located at the 2/3 of glissando for an infraliminal one. These results confirmed PP estimations which could be drawn from the work of Nabelek & al (1970). Moreover, Rossi pointed out that, even when the glissando is above the threshold, it is not perceived in its totality but only up to a point situated at approximately two thirds of the vowel. These data seem to imply that the mechanisms involved could be the same for pure tones and speech sounds. All the research mentioned above made use of linear frequency variations. This kind of variation, however, is rather different from the natural glissando of speech sounds. Furthermore, a linear variation does not distinguish between 2/3 of the duration and 2/3 of the frequency variation.

EXPERIMENTATION

In order to estimate the role played by the frequency variation and by the form of this variation, we synthesised frequency glissandos using a quadratic spline function, which provides a very close approximation to the 90 curves of natural speech (Hirst 1980). This technique makes it possible to vary the inflexion point of curves for glissandos with the same initial frequency, final frequency and duration and thus to dissociate variations of frequency and of duration.

**Stimuli**

Glissando:

3 glissandos using a vowel /a/
initial frequency: 100 Hz,
final frequency: 141 Hz (half an octave),
durations: 200 ms,
inflexion point located at 50 ms for G1, 100 ms for G2 and 150 ms for G3.

Comparison sounds:

vowel /a/ with fixed frequency and a duration of 100 ms. The frequency was modified from 100 to 150 Hz in 5Hz steps.

**Subjects**

18 subjects, with normal hearing.

**Experimental procedure**

The stimuli were presented by a constant method in 31 series divided into 3 experimental sessions. The first series was used for practice and is not included in the results.

**RESULTS**

For each glissando, the PP was calculated by a linear regression using the z transform. The experimental results (E) and PP predicted by the model 2/3 of the duration (D) and the model 2/3 of frequency variation (F) are displayed in table 1.

<table>
<thead>
<tr>
<th></th>
<th>E</th>
<th>D</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>120.4</td>
<td>134</td>
<td>127.3</td>
</tr>
<tr>
<td>G2</td>
<td>129.1</td>
<td>131</td>
<td>127.3</td>
</tr>
<tr>
<td>G3</td>
<td>127.7</td>
<td>123</td>
<td>127.3</td>
</tr>
</tbody>
</table>

The pitch differences are slight but significant at .04 (F(2,18) = 3.27).}

**DISCUSSION**

The comparison between experimental values and those predicted by the model show that the 2/3 of F model, give a fixed estimation of PP close to the observed values; while the 2/3 of D model correctly predicts the order of the experimental values. The model which predicts that the PP is near to the 2/3 of F is confirmed by our results; however this model cannot explain the differences between G1, G2 and G3. The pitch at 2/3 of the duration classifies the PP in the same order as the experimental values but does not provide an acceptable estimation of PP since the theoretical values are too far from the experimental values. It thus appears necessary to take into account both parameters, F and D.

If we assume that the auditory system integrates the final part of the glissando, a good enough approximation of PP is provided by the integration of frequency variation with respect to its duration (in fact the area beneath the curve). In order to obtain an estimation consistent with our results, however, we would need to assume that the integration applies not to a constant duration for the three stimuli, but rather to the last 120 ms for G1, the last 120 ms for G2 and only the last 100 ms for G1. Comparing these figures with the
glissando threshold curve, we note that this curve crosses the glissando rather close to these points. An attractive interpretation would consequently model the PP taking into account the glissando threshold. From our data we propose a model of glissando of this kind.

The pitch of glissando could result from a process of frequency integration from point where the signal is identified as glissando (tc = intersection of threshold curve and glissando) to the end of the signal. Thus:

\[
PP = \frac{1}{\int_{tc}^{tn} F(t) \, dt}
\]

\[PP \text{= pitch point}
F(t) \text{= frequency at the end of glissando}
Fc(t) \text{= Frequency at the point } tc\]

which at least in our experimental condition provides a fairly good estimation of the PP. Table 2 shows the values predicted by our model and the experimental values:

<table>
<thead>
<tr>
<th></th>
<th>PP Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>120.4</td>
</tr>
<tr>
<td>G2</td>
<td>129.2</td>
</tr>
<tr>
<td>G3</td>
<td>127.7</td>
</tr>
</tbody>
</table>

It should be noted that our model tends to overestimate the PP and that the difference between expected value and observed value is not negligible for G1. This could easily be corrected by modifying the glissando threshold function slightly or by adding an ad-hoc constant. We prefer to present here the values which result from the direct application of the model given above.

Further experimentation will be needed to determine the precise effect of the different parameters on the PP of glissando in order to improve the model.

Our results show that the PP of a glissando is a complex function of F and D. For the stimulus used in this experiment the PP is located in the vicinity of 2/3 of F, which is in agreement with data from Rossi (1971, 1978). Moreover, our results seem to show that the inflexion point of the glissando plays a slight but significant role in the perception of the PP.

REFERENCES

BRADY, P.T.; HOUSE, A.S.; STEVENS, K.K.
Perception of sounds characterized by a rapidly changing resonant frequency.

HIRST, D.J.
Un modèle de production de l’intonation.

NABELEV, I.U.; NABELEV, A.D.; HIRSCH, J.J.
Pitch of tone bursts of changing frequency.

NISHIMURA, Y.
Un modèle d’analyse automatique de la prosodie; accent et intonation en japonais.

ROSSI, M.
Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole.

ROSSI, M.
La perception des glissandos descendants dans les contours prosodiques.

TSUMURA, T.
Auditory thresholds of frequency change in brief tones.
SPEECH ANALYSIS OF ARABIC SOUNDS USING LPC METHOD

O. Abdel-Allm, A. El-Haddidi and M. Ezz-Elarab

Faculty of Engineering, Alexandria University, Alexandria, Egypt

INTRODUCTION

In classical Arabic there are some phonemes which have no exact equivalent phonemes in English. In order to get a perfect Arabic synthesizer, it is necessary and important to obtain the main acoustic parameters for each phonem in classical Arabic. This paper gives a study for the determination of the main acoustic parameters for emphatic and non-emphatic Arabic sounds. In many studies such as in /l/, /l/ and /l/, the stress was given on the effect of the emphatic process on the formant frequencies of the vowels when the consonant was pre- or post- this vowel. It is simple and easy to analyze the changes in the acoustic parameters of the vowels using the traditional methods such as the sonograms, but it is difficult to obtain the acoustic features of emphatic or non-emphatic sounds by using the same method. This paper will use linear prediction coding (LPC) method for the analysis of emphatic and non-emphatic Arabic sounds.

THEORY

In /l/ and /l/ it is given that the choice of an all pole model for the vocal tract can approximate the effect of antiresonances on the speech in the frequency range of interest, moreover, the location of poles is more important perceptually than the location of zeros. In the analysis of Arabic consonants we used an all pole model which is represented in Fig. 1. The output signal $S_n$ can be represented as a linear function of past outputs, present inputs and past inputs of $S$. Therefore:

$$ S_n = - \sum_{k=1}^{M} \alpha_k S_{n-k} + G \sum_{l=1}^{q} (E_l U_{n-l}) $$

where

- $\alpha_k$ = the predictor coefficients for the output $S_n$
- $b_l$ = the predictor coefficients for the input $U_n$
- $G$ = the gain factor
- $M$ = the order of the linear predictor (for past output samples)
- $q$ = the order of the linear predictor (for past input samples)

The control parameters that define the speech production model are:

1. The predictor coefficients of the time varying filter.
2. The gain factor.
3. The pitch period for voiced speech.

Assuming that the input ($U_n$) is totally unknown, then the signal ($S_n$) can be predicted with a certain approximation from a linearly-weighted summation of past samples with predictor coefficients.

If we have the set of predictor coefficients, then we can find the frequency response of the model for speech production. From this frequency response we can get the formant frequencies as well as the position of the peaks of energy for unvoiced sounds.

LPC ANALYSIS OF EMPHATIC AND NON-EMPHATIC SOUNDS

Speech signals are low-pass filtered (f_H = 10 KHz) and the output is sampled on a rate of (f_s = 20 KHz) and the sampled data is fed into a microcomputer to get the predictor coefficients as well as the spectral information for the inverse filter. For the analysis we used four consonant pairs (l/s(l/)/l/), (l/), (l/), (l/ and l/) and (l/). For each consonant we get four frames each of 10 m-sec interval. Fig. 2 shows two different waveforms for l/s and l/.

Table 1 gives an example for the obtained predictor coefficients and the gain factor for the sounds l/s and l/ for one frame.

Figure 3 gives the spectral information for l/s and l/.

DISCUSSION AND CONCLUSION: From the obtained results we can conclude the following:

1. Consonant l/s(l/)/l/ has maximum energy concentration in the range of (5351-5730 Hz) for all frames, while consonant l/s(l/)/l/ has a maximum energy concentration in the range of (3070-3516 Hz) as in Fig. 3.
2. Consonant l/ has maximum energy concentration in the range (646-820 Hz), while consonant l/ has maximum energy concentration in the range (351-460 Hz), as in Fig. 4.
3. Consonant l/ has maximum energy concentration in the range (429-540 Hz) while consonant l/ has maximum energy concentration in the range of (234-294 Hz), as in Fig. 5.
4. Consonant /k/ and /g/ have maximum energy concentration in the range of (1745-2050 Hz), while consonant /s/ has maximum energy concentration in the range of (618-859 Hz), as in Fig. 6.
5. The predictor coefficients for the emphatic and nonemphatic Arabic sounds are significantly different as shown in Table 1 for /s/ and /g/.
6. LPC analysis method can be used to get the speech parameters direct from the Arabic consonant sounds, and hence there will be no confusion in the meaning of the synthesized emphatic and nonemphatic Arabic sounds because they have different parameters.

REFERENCES
2. S.A. Adem, Report from Uppala University, Dept. of Linguistics, no. 11, 1983 ruul 11.
THE CLICKS IN ZULU: AN ARTICULATORY AND ACOUSTIC STUDY

A. Giannini, M. Pettorino and M. Toscano

Istituto Universitario Orientale, Fonetica Sperimentale, Napoli, Italia

From their beginnings phonetic researches have tried to segment the speech chain into articulatory, acoustical and functional units in order to classify them into well defined categories.

In this paper we'll not examine closely the problem whether it is possible to segment a continuum into a succession of discrete units or not. However we must recognize that it is certainly useful, in order to make a descriptive analysis of the speech, to create categories differing from one another in the presence or absence of a certain parameter.

These categories are often seen as universal rather than exclusively functional and this can give rise to many misunderstandings. In fact, once verified that a given articulation belongs to a category, one is often inclined to believe, on the basis of the binary organization, that a corresponding articulation must exist in the oppositional class. This happens even though one has already sufficient data to exclude a priori this possibility. The worst of it is that, as it often happens, when you look for something in the wrong place, you end by finding it.

In order to clarify what we have said above, it seems useful to examine articulations completely different from those for which traditional categories have been created.

In this study we refer to the Zulu clicks. The Zulu language belongs to the Nguni group of South-Eastern Bantu. The earliest works on Zulu go back to the close of the first half of XIX century. Since then many works appeared, the most valuable both in quality and quantity being the work of C.M. Doke.

According to Doke, Zulu speech sounds are divided into vowels and consonants, this last being subdivided into "plain consonants" and "click consonants".

Many experimental studies have examined clicks in order to identify their articulatory characteristics. It is generally accepted that clicks are produced by a particular articulatory mechanism called "ingressive velaric airstream mechanism". This mechanism consists of the cooccurrence of two points of closure of the tongue, the back one always being velar. The air enclosed between these two points of closure undergoes a rarefaction owing to backward and downward movement of the tongue. When the front closure is released the air rushes into the mouth to compensate the difference in pressure.

This kind of mechanism involves only the body of air that is in front of the velar closure, but this doesn't exclude the activity of the pulmonic airstream mechanism. In fact, behind the velar closure the air can flow through the nose when the soft palate is lowered or as Ladedegofed 1975 p.120 says: "Even if the soft palate is raised so that air cannot flow through the nose, the pulmonic airstream mechanism can still be used to keep the vocal cords vibrating for a short time during a click".

We have to underline that the temporal succession of the releases of the two closures is always clearly described, the first opening being the front one. On the contrary it is not sufficiently clarified the temporal succession of the two closings. In fact Doke 1926 p.124 affirms that "in preparation for the sounding of the clicks, the front of the tongue was first raised into position, the back of the tongue rising immediately after". Abercrombie 1966 p.31 and Catford 1977 p.72 talk about the 'velar initiatory closure'. Ladedegofed 1975 p.119 says that "at the beginning of this sound there are both dental and velar closures".

It is generally accepted both in traditional grammars and in phonetic experimental studies that in Zulu three different kinds of clicks exist: /c/ dental, /x/ dental lateral and /q/ post-alveolar. They can be produced in five different manners: voiceless, voiced, voiceless nasal, voiced nasal and voiceless aspirated.

The aim of this research is on one hand to clarify the articulatory mechanism of production of the clicks and, on the other hand, to examine spectrographic and electro-aerometric data, to verify whether the classification reflects a real articulatory dynamics or, on the contrary, it is the result of the tendency to arrange articulations into binary oppositional classes in spite of the experimental data.

For functional reasons we'll expose the experimental data following the traditional classification of the clicks.

PROCEDURE

For purposes of this experimental study we have two native speakers from Zululand read from a prepared list of words in which every kind of click appears in every phonological context. The words were recorded and analysed using a Voice Identification Series 700 Sound Spectrograph. For each word a broad band spectrogram was obtained. The fundamental frequency, air flow and intensity were measured using a FM 650, an Electro-Aerometer 510/4 and an IM 360 by F-J Electronic ApS. The duration of the steady state, the intensity of the release, the fundamental frequency, the oral and nasal air flow and the F1 and F2 loci of the preceding and following vowel have been analysed.

VOICELESS Clicks /c/ /q/ /x/

On the spectrogram the voiceless click is characterized by the absence of signal along the frequency scale followed by a burst of noise of strong intensity. On the spectrograms of /c/ and /x/ the releases of the two openings are almost always evident. The interval between the first and the second release is about 15 ms. On the contrary, as regards /q/ on the spectrogram only one release appears and we can suppose that the two openings are almost simultaneous.

In addition to this there is a noticeable difference in the distribution of the signal at the mo-
ment of the release. In fact for /c/ the signal is above 5000 cps, for /x/ above 1000 cps and for /q/ along the whole frequency scale. We have to notice that in almost all the cases there is a temporal interval between the release of the velar closure and the onset of the following vowel. Hefner 1952 p.119 hypothesizes that this delay is due to the presence of a glottal closure. In this case, however, the formants of the following vowel, contrary to what spectrograms clearly show, should start from their steady state level because during the glottal closure the vocal tract has already assumed the typical shape of the vowel. In our opinion the pause between the click and the following vowel must be explained considering the variations in pressure occurring inside the vocal tract. In fact the tracing of the air flow clearly shows that, at the opening of the front closure, there is an abrupt ingressive airstream and this necessarily involves an increase of the oral pressure. So, at the opening of the velar closure it is necessary a certain time in order to restore the difference of pressure between the inner and the outer air indispensible to create the egressive airstream.

The transitions of F1 and F2 of the vowels preceding and following the click have been calculated. With regard to the preceding vowel we have found that F1 locus is at about 250 cps and F2 locus is at about 2400 cps for all kinds of clicks. These data show that the first closure is the velar one contrary to the above cited Doke and Ladefoged's assertions.

We have examined F1 and F2 transitions of the vowel following /c/ /q/ and /x/. For all the three clicks F1 locus is at about 350 cps and F2 locus is at about 1000 cps. From an articulatory point of view these data confirm that there is a cooccurrence of two places of articulation, one velar and one in front of it.

As regards the manner of articulation the amount of ingressive air flow for /x/ is lower than that for /c/ and /q/. This difference in air flow means that in /x/ the opening of the front closure is not complete but it is partial and therefore the lateral nature of this articulation is confirmed.

VOICEDCLICKS /gc/ /qg/ /gx/

Neither the spectrograms nor the air flow tracing of /gc/ /qg/ and /gx/ shows any difference between the voiced clicks and their voiceless counterparts. We have to underline that during the steady state of voiced clicks there is neither nasal airstream nor glottal vibrations. In other words the so-called "voiced" click is in reality voiceless and this explains why many words can be written either using the "voiceless" or the "voiced" click indifferently without changing their meaning as, for instance, /qala/ and /qala/ (being on the watch), /eka/ and /geka/ (be enunciated).

VOICELESS AND VOICED NASAL CLICKS /nc/ /nq/ /nx/ /ngc/ /ngq/ /ngx/

Our experimental data show that the only difference between the nasal and oral clicks is given by the fact that, during the steady state of the nasal click, there is an escape of air through the nasal cavities with glottal vibrations.

We have to underline that the so-called "voiceless" nasal clicks are in reality voiced so, we can assert that in Zulu there is only one kind of nasal click, that is the voiced one. However, the vowels following either the "voiceless" or the "voiced" nasal click are substantially different. In fact the vowel occurring after the "voiceless" nasal click is nasalized whereas the vowel occurring after the "voiced"click is not nasalized, as the nasality ends at the opening of the front closure. This different behavior of the velum necessarily involves a different trend of F0. In fact in the "voiceless" nasal click the glottal vibrations continue until the vowel without variations while in the "voiced" ones there is a voiceless gap between the velar release and the vowel.

The fact that the vowel is either nasal or oral according to whether the preceding nasal click is either "voiceless" or "voiced" can be considered as a way of maintaining the functional opposition that otherwise would be lost. This is confirmed by the fact that words differing only in the "voiceless" and "voiced" nasal click maintain different meanings.

CONCLUSIONS

The data gathered in this experimental study on the Zulu clicks show that the click is the result of a velar ingressive airstream mechanism due to the cooccurrence of two oral closures one of them always being velar.

As regards the temporal sequence of this mechanism the velar point of articulation is the first to close and the front one is the first to open.

There are three different kinds of Zulu clicks: dental, dental lateral and alveolo-palatal. Each of them can be realized either voiceless or voiced nasal.

From a descriptive point of view these data suggest that the traditional classification of clicks that, as said above, recognizes the existence of three places of articulation combined with five manners of production, can be noticeably simplified.

In conclusion it seems to us that the Zulu clicks represent a significant example of the tendency to arrange the speech sounds into binary oppositional classes even if this is in evident contradiction with the experimental data.

REFERENCES


ACOUSTIC FEATURES OF INDONESIAN AND MALAY VOWELS

H. Nakashima, T. Kobayashi and O. Kakusho

Setsuman University,
17-8 Ikedanakamachi, Neyagawa, Osaka 572, Japan.

INTRODUCTION

Indonesian and Malay originate in the same language and have the almost same vowel systems. The first aim of this study is to extract the formant frequencies of Indonesian and Malay vowels, vis./i,e, a, o, u/ and a central vowel /e/ and to obtain the distributions of the formant frequencies for each vowel from the speech samples. Secondly, Indonesian is used as a common language in Indonesia, but there are many regional languages and most speakers of Indonesian have a knowledge of at least one regional language, such as Javanese. Therefore it is supposed that Indonesian is affected by the regional language. Another aim is to examine the effects of a regional language, Javanese, on Indonesian vowel system.

METHOD

Speech Samples

We selected Indonesian (Malay) words which have the six vowels in the initial position. There were 20 male adults (15 Indonesian and 5 Malay) taking part in the experiment. We divided Indonesian informants into three groups, the group A consists of the informants who speak Javanese in daily conversation and whose ages are from 20 to 23, the group B's informants also speak Javanese, but their ages are from 49 to 54, and the group C's informants speak Indonesian in daily conversation and their ages are 20 and 21. Group D's informants are Malay speakers and their ages are from 24 to 36. The informants were asked to read a set of words naturally. The recording were made in silent circumstance on a high quality tape-recorder.

The list of 26 Indonesian (Malay) words is as follows,

<table>
<thead>
<tr>
<th>Indonesian</th>
<th>Malay</th>
</tr>
</thead>
<tbody>
<tr>
<td>ibu</td>
<td>ube</td>
</tr>
<tr>
<td>(mother)</td>
<td>(four)</td>
</tr>
<tr>
<td>ikut</td>
<td>-</td>
</tr>
<tr>
<td>(follow)</td>
<td>(four)</td>
</tr>
<tr>
<td>ini</td>
<td>-</td>
</tr>
<tr>
<td>(this)</td>
<td>(four)</td>
</tr>
<tr>
<td>itu</td>
<td>-</td>
</tr>
<tr>
<td>(that)</td>
<td>(four)</td>
</tr>
<tr>
<td>isi</td>
<td>-</td>
</tr>
<tr>
<td>(contents)</td>
<td></td>
</tr>
<tr>
<td>enak</td>
<td>adil</td>
</tr>
<tr>
<td>ekor</td>
<td>ajar</td>
</tr>
<tr>
<td>ecer</td>
<td></td>
</tr>
<tr>
<td>erek</td>
<td></td>
</tr>
<tr>
<td>(nice)</td>
<td>(tail)</td>
</tr>
<tr>
<td>(tall)</td>
<td>(sell)</td>
</tr>
<tr>
<td>(sell)</td>
<td>(luggage)</td>
</tr>
<tr>
<td>(whats)</td>
<td>(to be)</td>
</tr>
<tr>
<td>apa</td>
<td>(just)</td>
</tr>
<tr>
<td>ada</td>
<td>(teach)</td>
</tr>
<tr>
<td>adil</td>
<td>(learn)</td>
</tr>
<tr>
<td>ajar</td>
<td>(learn)</td>
</tr>
<tr>
<td>(what)</td>
<td></td>
</tr>
<tr>
<td>obat</td>
<td>(medicine)</td>
</tr>
<tr>
<td>oleh</td>
<td>(by)</td>
</tr>
<tr>
<td>orang</td>
<td>(person)</td>
</tr>
<tr>
<td>chop</td>
<td>(guide)</td>
</tr>
<tr>
<td>umum</td>
<td>ubi</td>
</tr>
<tr>
<td>(general)</td>
<td>udara</td>
</tr>
<tr>
<td>(potato)</td>
<td>uji</td>
</tr>
<tr>
<td>(air)</td>
<td>ukur</td>
</tr>
<tr>
<td>(test)</td>
<td></td>
</tr>
<tr>
<td>(measure)</td>
<td></td>
</tr>
<tr>
<td>empat</td>
<td>enkeu</td>
</tr>
<tr>
<td>(four)</td>
<td>orang</td>
</tr>
<tr>
<td>enam</td>
<td>enam</td>
</tr>
<tr>
<td>(you)</td>
<td>(four)</td>
</tr>
<tr>
<td>(four)</td>
<td>(four)</td>
</tr>
<tr>
<td>(you)</td>
<td>(you)</td>
</tr>
<tr>
<td>(four)</td>
<td>(four)</td>
</tr>
<tr>
<td>(four)</td>
<td>(four)</td>
</tr>
<tr>
<td>(four)</td>
<td>(four)</td>
</tr>
<tr>
<td>(four)</td>
<td>(four)</td>
</tr>
</tbody>
</table>

Analysis

A total of 520 speech samples (26 words x 20 informants) were low-pass filtered to 4000 Hz and sampled at 10 kHz. These waveforms were analyzed by the autocorrelation method of linear prediction to estimate the formant frequencies within 20 msec frames, with the frames being advanced every 10 msec through the waveforms. Each frame was pre-emphasized by a first-order backward difference, multiplied by a Hamming window and used to compute a 12th order analysis filter. We obtained the formant trajectories over the waveforms and decided the first and second formant frequencies for each of the vowel tokens by averaging the formant values over the steady state portion of the vowels.

RESULTS AND DISCUSSION

The first and second formant frequencies, F1 and F2, which were extracted from the vowel tokens are described in the F1-F2 plane (see Fig. 1). The mean values and standard deviations of the extracted formant frequencies for each vowel per group of informants are shown in Table 1.

In Fig. 1, the distributions of F1 and F2 for /i/ and /u/ are very compact. The standard deviations of the formant values are small and we find little difference of the formant values among the groups in Table 1.
In the case of /a/, it seems that there is little difference of the formant values among the groups in Table 1. We minutely examine the formant values for /a/ per word. The means and standard deviations of F1 for /a/ per word in each group are shown in Table 1. In the results of Indonesian speakers, we find the differences of the means of F1 values for /a/ between in the word "obor" and in other three words, but there isn't such difference in Malay speakers. These differences for Indonesian speakers are greater in group A and B than in group C.

In the case of /e/, the distributions of the formant values for Malay speakers are compact, but those of the formant values for Indonesian speakers are diffuse and it seems that there are two clusters. In Table 1, we cannot find the difference of the means of the formant values among the groups of Indonesian speakers. The means and standard deviations of F1 for /e/ per word in each group are shown in Table 3, and it shows that there are the differences of the means of F1 values for /e/ between in the words "enak, ekor" and in the words "ecer, ecet". These differences are greater in group A and B than in group C.

<table>
<thead>
<tr>
<th>VOWEL</th>
<th>GROUP</th>
<th>F1</th>
<th>F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>A</td>
<td>350(26)</td>
<td>2402(181)</td>
</tr>
<tr>
<td>B</td>
<td>319(19)</td>
<td>2280(91)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>337(35)</td>
<td>2396(98)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>301(22)</td>
<td>2235(118)</td>
<td></td>
</tr>
<tr>
<td>/e/</td>
<td>A</td>
<td>612(109)</td>
<td>2046(256)</td>
</tr>
<tr>
<td>B</td>
<td>609(131)</td>
<td>1968(194)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>590(83)</td>
<td>2096(139)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>501(33)</td>
<td>2020(128)</td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>A</td>
<td>828(111)</td>
<td>1449(65)</td>
</tr>
<tr>
<td>B</td>
<td>813(76)</td>
<td>1468(47)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>852(48)</td>
<td>1425(79)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>726(89)</td>
<td>1281(240)</td>
<td></td>
</tr>
<tr>
<td>/o/</td>
<td>A</td>
<td>565(82)</td>
<td>943(78)</td>
</tr>
<tr>
<td>B</td>
<td>492(87)</td>
<td>803(69)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>538(69)</td>
<td>855(57)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>492(66)</td>
<td>849(116)</td>
<td></td>
</tr>
<tr>
<td>/u/</td>
<td>A</td>
<td>360(43)</td>
<td>788(78)</td>
</tr>
<tr>
<td>B</td>
<td>317(42)</td>
<td>639(35)</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>339(39)</td>
<td>780(30)</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>331(27)</td>
<td>758(64)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. The mean values and standard deviations in Hz of the formant frequencies extracted from the vowel tokens.

<table>
<thead>
<tr>
<th>obat</th>
<th>oheh</th>
<th>orang</th>
<th>obor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group A</td>
<td>535(50)</td>
<td>547(90)</td>
<td>531(64)</td>
</tr>
<tr>
<td>Group B</td>
<td>439(280)</td>
<td>441(39)</td>
<td>456(22)</td>
</tr>
<tr>
<td>Group C</td>
<td>516(27)</td>
<td>534(56)</td>
<td>518(67)</td>
</tr>
<tr>
<td>Group D</td>
<td>471(46)</td>
<td>480(38)</td>
<td>512(72)</td>
</tr>
</tbody>
</table>

Table 2. The mean values and standard deviations of F1(Hz) for /o/ per each word in each group.

In the case of /a/, it is shown that the standard deviation of F1 is large for Indonesian speakers, but on the contrary that of F2 is large for Malay speakers in Table 1. The means of F1 and F2 for Indonesian speakers are higher than those for Malay speakers.

Finally in the case of /a/, the distributions of the formant values for Malay speakers are rather compact, but those for Indonesian speakers are diffuse. In Table 1, we find the differences of the means of the formant values between group B and group A and C in Indonesian speakers. The 6 samples in 16 vowel tokens /a/ were pronounced as /u/ in group B (1/24 in group A and 2/20 in group C).

Indonesian vowels consist of 6 phonemes, but Javanese vowels consist of 8 phonemes, viz./i,e, e,a, o,u/ and a central vowel /a/. It is thought, therefore, that all Javanese and some Indonesian speakers pronounced Indonesian vowels and /e/ as /a/ or /o/ and /e/ or /e/, respectively.

CONCLUSION

We extracted the first and second formant frequencies of six Indonesian/Malay vowels and examined their distributions in F1-F2 plane. The results of this experiment indicate that there are phonetically different between in Indonesian and Malay vowel systems, and Indonesian vowel system is affected by the vowel system of speakers with the different regional background in Indonesia.

REFERENCES


MODELS OF FRICATIVE CONSONANTS INVOLVING SOUND GENERATION ALONG THE WALL OF A TUBE

Christine H. Shadle

Institute of Sound and Vibration Research, University of Southampton, Southampton S09 5NH, England

INTRODUCTION

Fricatives are produced when the vocal tract is constricted at some point sufficient to cause turbulence, which generates a hissing sound. Fant (1960) and Flanagan et al. (1975) have modeled fricatives by placing a shaped-noise pressure source downstream of the constriction in a transmission-line model. The source characteristics were varied until the synthesized speech sounded as natural as possible, or the synthesized spectra approached the closest to natural speech spectra.

In this 1995 study, experiments with mechanical models were discussed which were designed to test those fricative models more rigorously. In the simplest models used, consisting of a cylindrical constricted tube in a plastic tube, with or without a semicircular obstacle positioned downstream of the constriction, it was shown that the obstacle had the most marked effect on the spectral shape, and significantly increased the overall amplitude of the sound. Sound generation at the obstacle was shown to be well-modeled by a source pressure (a one-dimensional equivalent of flow dipole) presumed to be generated at the surface of any rigid surface contacted by turbulent air placed in a transmission line at the equivalent location of the obstacle. The shape and location of the obstacle was similar in shape to the teeth, known to be important in the production of /s, z/ (as in sin, shin). Spectra of mechanical models in which the placement of the obstacle and constriction were computed in teeth and tongue locations typical for /s, z/ were similar to human speech spectra of those fricatives. Thus, presumably, the pressure source at the location of the teeth is a good model for /s, z/.

For other fricatives the case is not so clear. The constriction is not positioned so as to direct the airflow onto the teeth; therefore there is less anatomical justification for a rigid obstacle normal to the flow. However, the no-obstacle case seems unlikely to be a good model because of its extremely low amplitude. We turn then to a comparison of speech and mechanical model data in search of the mechanism for fricatives other than /s, z/.

EXPERIMENT I

Method

Five subjects were recorded uttering the six unvoiced fricatives /θ, ð, ʃ, ʒ, s, z/ (as in sheon, fin, shin, sin, shin, ash), which varied in place of articulation from the lips to the velum. One subject also recorded /c/ (as in ich). The subjects were instructed to sustain each fricative for five seconds or more. The recordings were then analyzed using a spectrum analyzer to compute the averaged power spectrum.

Results

A major result of the speech analysis, which agrees with other such studies, is that while fricative spectra are quite consistent across tokens for a single speaker, there is considerable variation across speakers. In general, /s, z/ have the highest amplitude and are most similar in spectral shape across speakers. /θ, ð/ exhibit the next-highest overall amplitude and formant-like peaks which decrease in amplitude with frequency, as illustrated in Fig. 1. The frequencies of the peaks vary somewhat across speakers. /ʃ, ʒ/ are characterized by very low amplitude and great variation across speakers, so that it is difficult to describe a single typical spectral shape for each fricative.

Figure 1. Spectra of /θ, ð/ by female speaker EM. Solid lines = normal, dotted = intense production.

EXPERIMENT II

Method

The mechanical model experiments discussed in this paper were performed using a setup in which air passed from a pressurized air tank through a muffler and into the tube simulating the vocal tract. The sound generated was picked up by a microphone in the far field, and analyzed as in Experiment I.

The models were designed to mimic each fricative while still using simply geometric shapes for the outer tube and the constriction. The location of the constriction and its cross-sectional shape were chosen with reference to Fant's X-ray tracings of the vocal tract (1960). Perfect matches between speech and mechanical model spectra were not expected since the vocal tract shapes were still highly idealized.

The parameters used to compare fricatives and mechanical models were the overall amplitude, frequencies of significant peaks and troughs, and the dynamic range of the spectrum (the difference between the maximum and minimum amplitudes present across the entire frequency range). When an obstacle is introduced overall amplitude increases, troughs are evident, and as a result the dynamic range increases. A complete justification of these measures is given in Shadle (1989).

Results

The models used to mimic /ʃ, ʒ/ consisted of a constriction 1 cm in length and 0.08 cm² in area placed a distance 1/4 back from the mouth of a 17 cm plastic tube. The values of 1/4 were 4 and 6 cm, respectively, were used to imitate the different tongue positions for the two fricatives. Both a center circular constriction (CCC) and a flat-topped plug (a cylinder flattened along one side, FPT) that directed the air along the wall of the tube were tried. In either case, the front cavity resonances were excited, but the FPT generated more sound and thus the overall amplitude of the spectrum was higher. On this basis, and also in terms of the other spectral parameters, the flat-topped plug provided a better match for /ʃ, ʒ/. The difference in
amplitude between the spectra produced by those two constrictions became more pronounced, particularly at high frequencies, as the front cavity became shorter.

The fricatives /f, f, ʃ/ are all produced by constrictions formed at the teeth and lips. Therefore, there is almost no front cavity to speak of, and the acoustic radiation pattern (admittedly small, judging from the speech spectra) must be due to differences in constriction shape and perhaps area. In order to mimic these fricatives, therefore, a short front cavity 1 cm long was used in all cases, and several different shapes were used, including the CCC and the FTP. While it was possible to find a configuration that would match a production of each of the fricatives by at least one of the subjects in general, the constrictions mimicked the speech the best successfully for this group of fricatives. However, it was clear that the jet of air generated by the constriction needed to pass along some surface, presumably generating additional noise, in order to approach the levels and spectral shape found in speech.

Figure 2 shows spectra produced by the constricted circular constriction and the flat-topped plug for three values of $l_f$. Since the FTP spectra for $l_f = 4$ and $l_f = 6$ are similar to $\rho / S$ spectra, we would like to consider appropriate theoretical models of the sound generation process. It seems likely that dipole-like sound is generated by the air flowing along the wall of the tube; the greater efficiency of this mode of sound generation relative to the quadrupole sound present in a free jet (as for the CCC) accounts for the increased amplitude of the FTP relative to the CCC. However, this fails to account for the fact that the difference in amplitude increases as $l_f$ increases. Further work is needed to establish the critical parameters in sound generation and to develop appropriate source models for them.

CONCLUSION

It appears that all frication noise is dipole-like, but the dipoles are not necessarily modeled well by localized pressure sources since the surfaces at which they are generated are not always normal to the flow. For $l_f \geq 4$, where the flow is directed along the roof of the mouth, the sound that is propagated appears to be generated in the vicinity of the teeth and/or alveolar ridge, where there is evidence of flow separation. For the weak fricatives, further work is needed to establish the critical locations for sound generation and to develop appropriate source models for them.

Such a source model, however, focuses on the tube outlet, which is geometrically dissimilar to the lips. We must ask, therefore, whether the acoustic similarity of the FTP and the $\rho / S$ spectra is coincidental. To answer this question, a flow visualization study of the flow patterns in the various fricative configurations, using life-size mid sagittal cross-sections from Pant's X-ray tracing, was done. $\rho / S$ in particular showed clear evidence of flow separation at both the alveolar ridge and the upper teeth, approximately 1 cm downstream of the constriction. The FTP model is apparently correct in terms of the sound generation mechanism, but approximate regarding the location of the source. The good match with the speech data is probably due to the fact that X-rays of the subjects were not available, so that a 1 cm discrepancy in source location is not significant.

It is expected that all frication noise is dipole-like, but the dipoles are not necessarily modeled well by localized pressure sources since the surfaces at which they are generated are not always normal to the flow. For $l_f \geq 4$, where the flow is directed along the roof of the mouth, the sound that is propagated appears to be generated in the vicinity of the teeth and/or alveolar ridge, where there is evidence of flow separation. For the weak fricatives, further work is needed to establish the critical locations for sound generation and to develop appropriate source models for them. It should be noted that the standard cell function description of the vocal tract does not include the information necessary to choose a correct source model; nevertheless, a model with independent source and filter, and restricted to two-dimensional sound propagation, still seems at this juncture to be sufficient. This is based on work supported by an NSF-MDC 0002 grant, the Ford Foundation, and a grant awarded in 1984 by the North Atlantic Treaty Organisation.

REFERENCES


PALATOGRAPHIC AND ACOUSTIC MEASUREMENTS OF THE FRIкатIVE CONSONANT PAIR /s/ and /z/

M.H. Cohen and J.S. Perkell

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA

BACKGROUND

This study concentrates on the articulatory and acoustic contrasts between the coronal fricatives /s/ and /z/. Both /s/ to /z/ continuum and canonic /s/ and /z/ productions are investigated. Articulatory measures are derived from palatographic data and acoustic measures from spectrographic analysis. Stevens (1972) argues that the continuous movement of an articulator along an appropriate dimension during the production of a sound segment results in a change in the acoustic output that tends to be discontinuous. Regionally coarticulatory output defines acoustic correlates of a phonetic feature and are exploited in speech production and perception. The coronal fricatives /s/ and /z/ offer an ideal opportunity for investigating this theory. In a previous study on the production of the same fricatives, Perkell et al. (1979) observed a non-linear saturating relationship between articulatory and acoustic measures to accompany the production of the /s/ to /z/ continuum. Their articulatory measures were the making and breaking of contact between the inferior aspect of the tongue blade and the lower incisor, an essentially discontinuous measure. Shadle (1962) postulated mechanical tube equivalents to the vocal tract for the production of /s/ and /z/. Her models mimicked the differences in articulation as given by x-ray tracings (Fant 1960); namely, for /s/ the tongue tip is just behind the teeth and the tongue blade is held against the alveolar ridge forming a narrow constriction. A small cavity exists between tongue and teeth. For /z/, the tongue blade is raised forming a slightly broader constriction against the posterior part of the alveolar ridge. A significantly larger cavity is opened up between the tongue and the teeth. Shadle’s tube equivalents have four compartments; the back cavity, the constriction, the space between constriction and teeth, and the space between teeth and lips. There was good correspondence between spectra generated by the tube model and spectra from subjects. For /s/ and /z/, a turbulent noise source generated by the jet of air impinging on the obstacle (the teeth) is responsible for the excitation of the cavity between the constriction and the lips (front cavity). The difference in spectral shape between /s/ and /z/ is due to the sudden creation of a large space between the constriction and obstacle.

This study concentrates on the measurement of the in vivo positioning of the constriction with respect to the obstacle and the acoustic consequences of the creation of the sublingual cavity during productions of /s/ and /z/.

METHODS

Two male subjects KS and JP were used in this study. Each was fitted with a palatographic retainer holding an array of 64 contact points on the roof of the mouth (Rion DPO). Subjects were seated in a soundproof room. A two-channel recording was made with the acoustic signal in one channel and the amplitude modulated output of the recording adapter (Rion DPO) in the other. The subject was asked to produce several continuos of /s/ to /z/ by sweeping the blade of his tongue posteriorly from the alveolar dental ridge to as far back as possible, while minimizing lip protrusion. In some cases, the posterior configuration was more like retroflex /r/ than palatal /l/. Following this, canonic /s/ and /z/ were elicited from repetitions of /s/ and /z/ as well as of "she sells sea shells by the seashore." The two signals were then simultaneously digitized at 20 kHz. The palatographic information is demodulated and decoded so that the contact pattern of the tongue on the roof of the mouth can be displayed and plotted in 16 msec frames along with a time-aligned but downsampled acoustic signal. Figure 1a shows two palatographic frames during productions of canonic /s/ and /z/ from one subject. The contact is shown between the frontmost contact and the row having the narrowest opening is converted to distance and gives constriction-to-obstacle distance. This measure is tabulated for each 16-msec palatographic frame. Since the average distance between contacts is 5 mm, constriction location between contacts is interpolated assuming constant velocity of the constriction between adjacent contacts. Spectral analysis is carried out on the pre-emphasized acoustic signal. Log magnitude spectra are computed over 32-msec windows (2 palatographic frames) for the duration of the utterance. Cepstral smoothing is applied and the frequency of the highest spectral amplitude is automatically calculated. Figure 1b shows the log-magnitude and cepstrally smoothed spectra for the canonic /s/ and /z/ in Fig. 1a.

![Fig. 1: a) Palatographic imprints of the tongue on the roof of the mouth for canonic /s/ and /z/. (Subj. KS). b) Log-magnitude (shifted -10dB) and cepstrally smoothed spectra for /s/ and /z/ above.](image)

RESULTS

Both subjects were able to produce the continuum, but of the two subjects, KS was the most able to execute the manoeuvre more slowly and consistently. Figure 2 shows a plot of the frequency of major excitation as a function of constriction/obstacle distance for subject KS. There is a region of constriction location, between 2 and 6 mm, over which the frequency of major excitation doesn't change. Abruptly, at 6 mm, the frequency of major excitation drops precipitously from around 5 kHz to 2.3 kHz for approximately a 2-mm change in constriction/obstacle distance. This must
correspond to the sudden opening up of the sublingual cavity and the resulting large percentage increase in volume of the front cavity. Beyond a constriction/obstacle separation of 9 mm, the change in frequency with distance is more gradual suggesting a linear increase in the volume of the front cavity with posterior movement of the constriction. For subject JP, the transition from high- to low-frequency excitation at a constriction/obstacle distance of about 12 mm occurs over the same range of movement (2 mm). The dotted vertical lines on Fig. 2 show constriction on baseplate distances for canonical /a/ and /æ/ productions by subject KS. Constriction widths and lengths for the canonical /a/ and /æ/ are, respectively, 9.5 and 5 mm (for /a/) and 14.5 and 7.5 mm (for /æ/) (subject KS); 13 and 5 mm (for /a/) and 14 and 12 mm (for /æ/) (subject JP). Assuming that the front cavity behaves like a uniform tube, we can calculate the equivalent length of this tube from the quarter wavelength of the frequency of excitation. Figure 3 shows the calculated front cavity length on the ordinate plotted against the measured front cavity length on the abscissa. The measured length is derived from the constriction/obstacle distance plus the obstacle/lip distance for subject KS (12 mm). The figure shows three regions: I, where the assumption of a uniform tube is correct (slope=1); /æ/ is produced in this region. II, where the front cavity cannot be described by a uniform tube because presumably the percentage increase in volume of the front cavity is abrupt when the sublingual space opens up. This corresponds to the region of acoustic "instability" in Fig. 2. III, where once again the front cavity behaves as a uniform tube (slope=1) and the length of the front cavity is determined by the sublingual space. /æ/ is produced in this region.

CONCLUSIONS

The results presented here largely corroborate the findings of Perkell et al. (1979) and Shadle (1985). Measurements taken from two subjects give quantitative data on the constriction/obstacle distance and the width and length of constrictions for productions of /a/ and /æ/. The findings from both experimental subjects fit well with those of Stevens (1972) on the quantal nature of speech. The acoustic "plateaus" along which /a/ and /æ/ occur are separated by a region of rapid acoustic change. The extent of change, evidenced by the drop in the frequency of maximum excitation from 5 to 2.3 kHz, is caused by the sudden creation of a sublingual space over a very small range of posterior movement of the constriction. This twofold change in frequency heightens the perceptual difference between /a/ and /æ/.

Future work will concentrate on estimating the volume of the front cavity as well as the cross-sectional area and volume of the constriction using dental impressions and x-ray images.

[Supported in part by an NIH grant. Thanks to Ken Morey for software for the electropalatograph and to Ken Stevens for many useful discussions.]

REFERENCES

PRELIMINARY SUPPORT FOR A "HYBRID MODEL" OF
ANTICIPATORY COARTICULATION

J.S. Perkell and C-M. Chiang

Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139,
U.S.A.

BACKGROUND

According to inferences (Perkell, in press) made from the "look-ahead model" of anticipatory coarticulation (cf. Henke, 1967), a single, smooth gesture toward an articulatory target should begin as soon as there are sufficient, prior requirements on the articulator. Alternatively, the "temporal model of speech production" (Bell-Breri and Harris, 1981) predicts that there are no such context-dependent adjustments of movement onsets and at a given rate, the anticipatory gesture always begins at the same time (i.e., it is "time-locked"—c.f. Sussman and Westbury, 1981) with respect to the (acoustic) onset of the target segment. In this framework, conflicting requirements of successive phonetic segments on the articulator presumably are resolved by "co-production" (Fowler, 1980), which entails an overlapping and summation of such requirements.

One plausible accounting of the differences between the two theories and supporting sets of data comes from Bladon and Al-Barenji's (1962) finding of two different patterns of anticipatory velar movement for nasal consonants: a single, smooth opening gesture and a two-stage gesture. Both the single-stage gesture and the initial gradual stage of the two-stage gesture had onsets that occurred as soon as "permitted" by the termination of a preceding obstreperous consonant. When it occurred, the onset of the second, higher velocity stage of the two-stage gesture coincided with the oral closure gesture for the nasal consonant. Both patterns occurred across speakers of different languages, and across multiple repetitions of the same utterances by individual speakers. Bladon and Al-Barenji speculate that findings supporting look-ahead arise from observations of initial onsets of one or the other gesture type, and findings supporting time locking arise from observations of the onsets of the second, higher velocity component in two-stage gestures.

METHODS

This paper presents observations of lip protrusion movements from one speaker of American English that agree with certain of Bladon and Al-Barenji's observations of velar movements. In this experiment, protrusive movements of the upper lip were recorded with a strain-gauge movement transducer, along with the acoustic signal. The subject spoke pairs of one-syllable words in the carrier phrase "its a ________ again, with stress on the second word of the pair. The words were chosen so that the target vowel /u/ (or control vowel /i/) in preceded by a VCV or CVV string, where V is /i/ or /a/, and CV are combinations of subsets of /k,t,s,n,h/. Example "target" and control pairs of words are "leak hoot" and "leak heat". There were 29 target and 29 control utterances spoken 15 times in random order.

Signals were digitized simultaneously at two different rates and processed into two time-synchronized signal streams. One signal stream contains the high-bandwidth (2012 Hz) speech signal and the other contains the low-bandwidth (256 Hz) lip protrusion signal, along with velocity and acceleration signals and algorithmically-determined marks at times of acceleration peaks. Acoustic segment boundaries in the speech signal were labeled with interactive audition and graphics. These data and signals were used to automatically generate a "sequence plot" for each utterance. Figures 1-4 show such sequence plots, each containing displacement versus time traces for the first 10 examples of one utterance. The line-up time, indicated by the vertical line is at the time of onset of voicing for the "target" vowel, /u/. The symbols on the plots indicate onset and offset of the protrusion gesture as determined by the beginning and end of positive velocity (boxes), times of acoustic events (triangles), the first of which is at the time of offset of regular glottal vibration for the preceding unvoiced vowel, and time of largest acceleration (vertical tick mark). These figures illustrate observations which are representative of the entire data base.
RESULTS

Figures 1 and 2 are for the utterances "leak hoot" and "leaked coot" respectively. Protrusion onset (first box) is gradual and approximately coincident with the end of the /i/. The largest acceleration peak (tick mark) is approximately coincident with the second triangle, (/k/ release in "leak hoot" and /t/ release in "leaked coot"). The major acceleration peak marks the onset of the larger, second component of the protrusion gesture, which is approximately similar in overall shape for the two utterances. For both utterances, the protrusion maximum (end of the gesture) occurs about at the beginning of the vowel /u/. Figure 3 is a sequence plot for "lee neat". For almost all tokens presented in this movement onset which is approximately coincident with the end of the /i/ and a later, more pronounced movement component with onset marked by the major acceleration peak. In comparison with the traces in Figs. 1 and 2, the major acceleration peak in Fig. 3 appears to be more variably timed and occurs closer to the line-up point, and the protrusion maximum is more coincident with the end of the /u/ (rightmost triangle) than with the beginning. Figure 4 is a sequence plot for "lee suit". In this figure, there is more variability of the onset of movement with respect to the end of the /i/, but it almost always begins before the end of the /i/. There is also considerable variability of the timing of the major acceleration peak and overall shape of the gesture. For some tokens (1-4, 10) the lips remain largely unprotruded during most of the /s/ and in others (5-7, 9), significant protrusion takes place earlier, during or before the /s/. Variation between these two patterns across tokens also occurs for the traces in Fig. 3. In all four figures, the traces are nearly flat up to the end of the /i/ (first triangle), suggesting that there is no active retraction associated with the /i/. The traces (not shown) for the corresponding control utterances ("leak heat", "leaked kest", "lee neat", "lee seat") are virtually flat, indicating that there are no protrusion "goals" for the consonants in an /i/-/i/ environment.

DISCUSSION

In general, the results suggest that 1) there are interactions between the vowel /u/ and different preceding consonants which result in protrusion gestures that have different temporal and spatial characteristics, 2) the onset of the overall protrusion gesture is linked to the end of the preceding unrounded vowel, 3) the overall gesture is "partitioned" around the time of the major acceleration peak into a gradual first component and a more rapid second component, 4) the exact nature of the partitioning is dependent on the "importance" of these requirements, the partitioned gestures can be variable in appearance from one token to the next.
VOISEMENT ET ASSIBILATION DES OCCLUSIVES "SONORES" D'ENFANTS QUEBECOIS (ETUDE ELECTROGLOTTOGRAPHIQUE)

Jean-Pierre Goudailier
Laboratoire de Phonétique, U.F.R. de Linguistique, Université René Descartes, Paris, France

Faisant suite à un ensemble de recherches que j'ai menées au sujet des occlusives françaises d'enfants francophones de 6-10 ans (J.-P. Goudailier, 1981, 1983, 1985), la présente analyse est tirée de l'étude des prononciations (environ 2050 occlusives "sourdes" et "sonores") relevées dans une population de 20 enfants québécois. Il s'agit d'une enquête effectuée dans le cadre du projet intégré "systèmes phonologiques d'enfants scolarisés francophones québécois et québécois âgés de 6-10 ans" associant les universités de Laval (Québec) et de Paris, avec des parents de l'Université de Paris, Diderot. La méthode instrumentale (électroglottographie) est la même que celle exposée dans les publications antérieures (cf., entre autres, J.-P. Goudailier, 1983, p. 267). Les mesures ont été faites à partir de tracés occlusifs (tracé oscillo et électroglottographique EGG) (cf. figures 1 à 9)M.

Les données relatives au corpus québécois concernent deux sujets qui avaient été relevées par les études précédentes, plus particulièrement en ce qui concerne l'âge des enfants. Ceci est confirmé par le manque d'occlusives phonologiquement "sonores" : a) les occlusives vélaire présente une plus grande tendance à se dévoiler que les dentales, qui tendent, elle-mêmes, à être plus dévoilées que les labiales (J.-P. Goudailier, 1981, p. 267); b) si l'on compare les positions initiales absolues de mot, initiale non absolue (mot précédé d'un article) et intervocalique, on voit que les quatre types de réalisations sont une valeur d'indexation régionale.

Le classement des occlusives phonologiquement "sonores" en termes de nombre d'apparition des vibrations laryngées (M.A.V.) et de prononciation par le sujet (J.-P. Goudailier, 1985, p. 48), c) le nombre d'articulations d'âge et d'âge d'âge (J.-P. Goudailier, 1985, p. 48); d) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); e) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); f) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); g) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); h) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); i) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); j) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); k) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); l) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); m) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); n) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); o) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); p) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); q) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); r) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); s) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); t) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); u) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); v) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); w) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); x) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); y) la durée d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48); z) le niveau d'articulations d'âge d'âge (J.-P. Goudailier, 1985, p. 48).

Tout comme pour les prononciations françaises, les occlusives "sourdes" québécoises présentent un M. A.V. de type ou de type. Ce dernier type correspond aux occlusives aspirées (V.O.T. < 35ms) et non aspirées (V.O.T. > 35ms). Le [g(h)] de la fig. 4 est de type 4 (V.O.T. = 42ms; occlusion = 100%) et le [p(h)] de la fig. 5 de type 5 (V.O.T. = 55ms; occlusion = 120%).

Pour des mots tels /lét/, /etl/, /tign/, /ädjë/ /nadj/, /kmodil/, /fodjem/, /diàv/, etc. ou pour le prononcés personnel /ty/, /l'article partitif /dy/, etc. le corpus québécois montre toute son originalité par rapport aux autres langues. En effet, ainsi que j'ai communément reconnu pour le franco-canadien et le franco-ontarien (cf., entre autres, P. Léon, 1979; A. Marchal, 1980, A. Thomas, 1985), les dentales /t/ et /d/, lorsqu'elles se trouvent devant une voyelle haute /i/ ou /y/, comprennent une phase d'assibilation qui correspond sur le plan acoustique à une assibilation. Ceci est constaté chez tous les enfants québécois analysés ; la fig. 7 présente un exemple relatif à l'occlusive "sourde" [] de /i/ et /y/ réalisée [t/] avec une fréquence de 75ms (occlusion = 100%). En termes de M.A.V.L., le type 7 est attribué à de telles articulations. Par ailleurs, une relation évidente peut être mise au jour entre l'assibilation et l'absence de stress. Cela est illusoire que le [d'] de la fig. 8, qui a une phase d'assibilation de 45ms ne comporte des vibrations que pendant 5ms au début de son articulation : plus de 50% de la phase d'assibilation ne comprendra donc aucune vibration des cordes vocales (70ms/125ms). Il s'agit ici d'un M.A.V.T. de type 7/7. Le type 2/7, pour lequel l'assibilation n'est jamais pas de délai important, est plus rare (1% du total). A la fig. 9, le [d'] a une phase d'assibilation de 35ms et une occlusion sonore de 75ms ; cependant, l'amplitude des vibrations décroit sur le tracé EGG pendant la phase d'assibilation.

Afin de rendre compte des faits québécois, il convient donc d'ajouter au classement en termes de M.A.V. le type 7 (assibilation) pour les occlusives "sourdes" (type 3/7 (assibilation avec voilement) et 2/7 (assibilation sans voilement)) et Monoarticulaires la phase d'assibilation "sonores" et de constater, de point de vue articulaire et acoustique, que l'assibilation assi- bilation occasionne une désorisation importante de l'occlusion des consonnes dentales "sonores".

Références bibliographiques
Goudailier Jean-Pierre, Exemple de traitement de l'opposition de "sonorité" par des enfants de Cours-Prépa-
Goudailier Jean-Pierre, Diversité des possibilités de ma-
etalisation du traitement de voilement : étude électro-
Léon Pierre, Contribution aux études de phonétique au Canada, Linguistique expérimentale et appliquée au Ca-
da, Montréal, Didier, 1979, p. 59-132.
Marchot Alain, L'articulation de [i] et [a] en français de Montréal, Travail de l'Institut de Phonétique d'Al-
ternance, 1980, p. 7-95.
Thomas Alain, L'assibilation en franco-ontarien, Infor-
mation & Communication (Toronto), 4, 1985, p. 65-80.
fig. 1 : [ɔheː] (M.A.V.L. 1); fig. 2 : [(ɔ)g(ŋ)] (M.A.V.L. 2); fig. 3 : [ɔŋ(ŋ)] (M.A.V.L. 3); fig. 4 : [b(ŋ)] (M.A.V.L. 4); fig. 5 : [ɔŋ] (M.A. V.L. 5).

fig. 6 : [ɔpə] (M.A.V.L. 6); fig. 7 : [ɪŋ] (M. A.V.L. 7); fig. 8 : [məd] (M.A.V.L. 3/7); fig. 9 : [ŋd] (dans (k(ŋ)xɔ(xₕ)xₙ)) (M.A.V.L. 3/7).

Remerciements : À Pierre Martin et Conrad Oullon du Département des Langues et Linguistique de l'Université de Laval (Québec) pour leur aide.

En France, cette recherche a bénéficié de subventions des Ministères de l'Education Nationale et des Relations extérieures (projet de coopération 01.01.138); au Québec, il s'est agi de subventions des Ministères de l'Education et des Relations internationales (projet de coopération 07.01.138).
THE NASOPHARYNGEAL TRACT: A TARGET FOR NASALITY.
ACOUSTIC SIMULATIONS VS. SWEEP TONE MEASUREMENTS.

G. FENG, C. ABRY & B. GUÉRIN
Institut de la Communication Parlée
46, avenue Félix Viallet - 38031 Grenoble Cedex

RECALL
In a preceding study (FENG et al., 1985) we proposed that all nasal vowels should be considered as dynamic trends towards a consonantal articulatory target, i.e. the nasopharyngeal tract (the nasal tract plus the pharynx), which offers a relatively stable acoustic structure, slightly sensitive to vowel variations, contrasting with the well-known great plasticity of the oropharyngeal one.

Continuing on a suggestion by MAEDA (1984), we found out that the main acoustic characteristics of our nasopharyngeal tracts — simulated for all French vowels — could fill the space left by articulatory possibilities of the oropharyngeal one, namely a "gap" between [u] and [u], in the F1-F2 plane, as evidenced by MAEDA's articulatory modeling (MAEDA, 1984) and phonological inventories. The question we face now is: what are the main solutions available in articulatory-acoustic simulations to obtain such a target for all vowels? And more interesting, could one evaluate them with respect to their coherence with sweep-tone data for nasals?

1. THE NASOPHARYNGEAL TARGET.

The idea of conceiving nasal vowels as being related to consonants is as old as evidence obtained from phonological processes, mainly historical ones: nasal vowel systems originate essentially from assimilation of oral vowels by adjacent nasal consonants (review in PERNY, 1965 & 1972). The nasopharyngeal target is supported as follows. Consonantal place contrasts neutralize into the so-called nasal appendix (g-like) — explained articulatorily by the lowering of the velum towards the tongue thus tending to block the oral port. And this occurs before consonant deletion. But even after complete deletion, there remains in phonetic realizations of proper nasal vowels a somehow evolutive (if not diptongal) character, due to the sluggishness of the velum lowering as shown by several authors (see LINDHOLST, 1971, for French).

Simulations of nasopharyngeal tracts with different vowel configurations — for their pharyngeal part — are scarce (HERLIER, 1984, with 4 French vowels). However, they offer, from our point of view, a lot of advantages (FENG et al., 1985).

But the only point focused on in this paper is the striking acoustic similarity between nasopharyngeal targets for all vowels. We will illustrate this issue with simulations for French vowels (FENG et al., 1985). The still recognizable structure of oral vowels shifts for nasopharyngeal appendices in a small 300-1000 Hz region (Fig. 1).

2. NASOPHARYNGEAL AND NASAL TRACTS. SIMULATED TRANSFER FUNCTIONS AND SWEEP TONE RESPONSES.

It's now appropriate to show that such a result can be obtained with three available solutions.

They are essentially proposals to obtain a sufficiently low peak (say 300 Hz) — an apparently important cue — for nasal vowels, which is particularly difficult to simulate for low vowels (like [e]).

Let's call them for convenience: the small nostrils, the sinuses (maxillaries) and the glottal forms solutions. The latter proposal (MAEDA, 1984) will not be considered here as we shall only focus on the coherence of transfer function calculations with sweep tone data.

2.1. Small nostrils.

It's easy to show that a large nostril output (say 2 cm² equivalent) brightens the first pole, shifting the nasopharyngeal zone to the center of the acoustic space (Fig. 2). We need in our case (Fig. 1) a 0.6 cm² equivalence.

The realism of such a solution is difficult to estimate. Without any statistical data on the nasal-tract, values given in the literature vary from 1 to 10. Instead of 0.2 cm² (HOUSE & STEVENS, 1956) or 2 cm² (MAEDA, 1982) or 1.5 cm² (B. JUODIS & PANT, 1964) we chose as acoustic equivalence 0.6 cm², a value which may occur at limen nasii (the most constrained passage in the nostrils, PALACIOS & al., 1980, GA-12-A). Our choice was guided by the fact that acoustic sensitivity to anatomic variations decreases slowly up to 0.5 cm² which seems to be a "boundary". But to make a realistic decision in these part of our modeling is not simply a matter of anatomic statistical data viewed from their "quantum" (STEVENS, 1972) consequences on the acoustic level; it demands further studies on output equivalence (see such difficulties for lips).

2.2. The sinus system.

Recently, PANT (1983) used another alternative. Simulating the nasopharyngeal tract, he introduced an acoustic equivalent of the sinuses maxillaries (the biggest ones). With a simulated Helmholtz resonator tuned to precisely 399 Hz which gives the zero in the transfer function of the nasopharyngeal tract — he obtained a first peak sufficiently low (about 300 Hz).

Such a sinus simulated on a line analog (with losses) demands approximately 45 cm³ in volume.

Does this correspond to some anatomical reality? Others (e.g. MAEDA, 1982) use a 20 cm³ equivalent. Here too the anatomical data allow a good range of variability. AUST & DREITZER (1975) for instance, give a 1 to 7 ratio for the volume. The variability of the ostium (whose values are entered into the simulations as section and length) is even greater: from 1 to 20. Moreover, its liability between individuals (seasons, etc.) adds to this high variation (acoustically from 200 to 300 Hz) between individuals, according to LINDHOLST-GAUFFIN & SUNDRENG, 1976, p. 166). We must note that the sinus maxillaries are late in ontogenetic growth (after acquisition of a possible nasal vowel system).

From this anatomical point of view, the volume value given by PANT may seem high, but not completely unrealistic. And the only firm point we can mention in this anatomical competition between nostrils and sinuses is that the former are evidently less variable — and "more evenly shared" — than the latter.
So we should count on other types of data to confront the two solutions.

2.3. Looking at sweep tone measurements.

Sweep tone data have been obtained since 1964 by FUJIMURA & LINDQVIST, for the vocal tract, with some nasalized sounds, including nasal vowels and consonants. Later LINDQVIST-GAUFFIN & SUNDBERG (1976) measured nasal tract responses. They tried to model with their data the presence/absence of sinuses, asymmetry, etc. Regarding the sinus maxillaries, the result was a first peak at 450 Hz in the nasal tract transfer function (650 Hz without this sinus).

MAEDA's simulation (1982) with a 20 cm$^3$ sinus corresponds to these values.

The question now is: could these values simulate correctly other sweep tone data available for the nasopharyngeal tract? FUJIMURA & LINDQVIST (1984) provided plausible values for the nasal consonant [ŋ] with [ŋ], around 250–300 Hz. We have shown elsewhere (FENG & al., 1985, Fig. 1B) that this first peak obtained with MAEDA's conditions is too high (300 Hz) in the nasopharyngeal case.

The same question could be asked conversely for FANT's conditions. His own simulation for the nasal tract sinus gives a first peak (297 Hz) too low compared with the 450 Hz indicated by LINDQVIST-GAUFFIN & SUNDBERG (1976, Fig. 2b).

From these results, we can see that, if one uses the sinus in simulating the nasal tract to obtain the first low peak, one would get no big differences (almost the same peaks) for the nasopharyngeal tract. The reason is clear: the zero given by the sinus being established, the first peak to the left of the zero will not change much whether the pharynx is connected or not.

This situation is not reasonable. And, moreover, it is not coherent with sweep tone measurements available for these two cases: they clearly show very different peaks. From the fact that a connection with the pharynx causes the first nasal tract peak to decrease, we posit that this first peak of the nasal tract alone is likely caused by a resonance frequency than by a sinus.

Our 0.6 cm$^3$ output equivalence solution offers the possibility to cover the distance between both the nasal tract and nasopharyngeal sweep tone measurements, with typical values of respectively 445 Hz and 307 Hz (Fig. 2a et b).

DISCUSSION AND CONCLUSION

The fact that sinuses offer less possibilities to maneuver in simulations to meet sweep tone data does not disqualify them completely from the nasality problem. They may contribute—if not crucially to the location of the lower peak—to other attributes of nasality. For instance, they add "flatness" (MAEDA, 1982-1983) in the spectrum, a flatness due to the presence of secondary poles around the so-called oral formants (see Fig. 4a et b), with the filling of the valley and the broadening of the peaks (STEVENS, FANT & HAWKINS, 1985). In this perspective the combination of the transfer function characteristics with the glottal formant (MAEDA, 1984) is an additional opportunity to explain the complexity of the real spectra.

**REFERENCES**


ACKNOWLEDGMENTS

S. MAEDA, G. FANT & L.J. BOÊ for discussions and suggestions; F. CHARPENTIER for software; R. SOKH for translation; D. VILLET for preparation of the manuscript.
ESTIMATION OF GLOTTAL FLOW FROM PRESSURE GRADIENT MEASUREMENTS IN THE TRACHEA AND THE PHARYNX

R. Cramer and L. Ruvos

Inst. of Phonetics, Nijmegen University, PO Box 9103, NL-6500 HD Nijmegen, The Netherlands.

INTRODUCTION

A complete characterization of the local acoustic behavior of a fluid requires that both pressure and flow are specified simultaneously. Consequently, if the development of more refined models of the human voice source is to be guided by knowledge about the actual behavior of the air particles, it is essential to have available techniques with which pressure and flow in the neighborhood of the glottis can be measured simultaneously.

Obtaining calibrated pressure recordings with an acceptable frequency resolution (> 5 kHz) is not much of a problem anymore. Catheters in which miniature, superconductor pressure transducers are mounted are readily available. Practical measurement techniques which allow for the recording of local, glottal air flow during normal speech production, however, are less easy to find.

In speech research there are three, more or less well-known techniques for measuring the glottal flow. The most commonly used technique is to estimate the glottal flow waveform by means of inverse filtering. With this technique not the glottal flow, but the flow at the mouth is measured. Assuming that the vocal tract acts as a linear filter, the glottal flow waveform is then estimated by (interactively) removing the effect of the filter. Obviously this method is very indirect and leaves heavily upon the assumption that the response of the vocal tract and/or some basic characteristics of the glottal flow waveform are known.

A second and much more direct way of measuring flow is hot-wire anemometry. By reducing the mass of the wire a frequency response up to 50 kHz can be obtained. However, because a hot-wire anemometer is insensitive to the direction of flow and also becomes rather unreliable at very low flow rates, it cannot be used for the investigation of acoustic phenomena in a straightforward way. By means of a setup in which a known air flow component is added to the one to be measured, these drawbacks can be overcome. At this moment, however, we do not know of any commercially available instrument which is small, robust and safe enough to be inserted via the nasal passage for making local flow measurements in the pharynx and/or trachea during normal speech production.

The third measurement technique uses the relation between pressure gradient and particle velocity in a fluid where plane wave propagation can be assumed. Because the assumption of plane wave propagation does not necessarily hold, however, the outcomes of this measurement technique must be interpreted with caution. An advantage is that pressure gradient can be measured in the pharynx and/or trachea by means of commercially available instruments.

In this paper an experiment is described in which pressure gradient signals were measured in the trachea and in the pharynx during normal speech production. From both the subglottal and supraglottal gradient signals separate glottal flow estimates can be derived. If the pressure gradient method is valid, these estimates should yield identical results irrespective of the vocal tract configuration. This appears not to be the case. By means of a critical comparison of the glottal flow estimates which are derived from subglottal and supraglottal measurements during the production of the vowels /a/, /u/, and /I/, possible causes for the discrepancies are sought.

MEASUREMENTS

Two Dutch adult males without any known vocal pathology read lists of VCV and CVVC words. During the reading the following signals were simultaneously recorded using a multi-channel FM-recorder:

1. photoglottogram (PGG) 2. electroglottogram (EGG) 3. subglottal pressure at 7.5 cm (P₁) and 2.5 cm (P₂) below the glottis 5. supraglottal pressure at 2.5 cm (P₃) and 7.5 cm (P₄) above the glottis 7. the acoustic speech wave at ca. 10 cm from the lips.

The recordings were subsequently digitized at an effective sampling rate of 10 kHz per signal and 12 bit amplitude resolution. From the digitized recordings only the /a/, /u/ and /I/ vocal portions were selected for further analysis. The wide band pressure recordings were made by means of a Millar PC-784(K) catheter, equipped with four miniature pressure transducers located at the distal end of the catheter and at 5, 10, and 15 cm from that end. The opening and closing moments of the glottis were determined by peak-picking in the (smoothed) first derivative of the EGG.

ESTIMATING GLOTTAL FLOW FROM PRESSURE GRADIENT

It is generally believed that during voicing the periodically interrupted glottal flow excites standing wave patterns in both the subglottal and supraglottal cavities. Moreover, because the dimensions of the trachea and the vocal tract are perpendicular to the direction of the flow, they are small when compared with the wavelengths of interest, it may safely be assumed that there occurs plane wave propagation. Therefore, the well-known relation

\[
\frac{\partial p}{\partial x} = -\alpha \frac{\partial \dot{v}}{\partial t}
\]

should hold. Here \(\frac{\partial p}{\partial x}\) is the pressure gradient in the direction of the flow, \(\frac{\partial \dot{v}}{\partial t}\) the time derivative of the particle velocity, and \(\alpha\) the mass density of the air at rest. Defining the position of the glottis as \(x = 0\), the approximate \(\frac{\partial p}{\partial x}\) at points 5 cm below and 5 cm above the glottis resp. by writing:

\[
\frac{\partial p}{\partial x} \bigg|_{x=-5} = (P₁(t) - P₂(t))/5
\]

\[
\frac{\partial p}{\partial x} \bigg|_{x=+5} = (P₃(t) - P₄(t))/5
\]

Fig. 1: Electrical equivalent of glottal region

From eq. (1) it follows that integrating the pressure gradient with respect to time and multiplying by the cross-sectional area of the tube yields the volume velocity. Thus, using eqs. (2)

we are able to estimate the volume flow at two locations positioned symmetrically around the glottis. An electrical equivalent circuit of the
measurement situation is given in Fig. 1. This figure also suggests a way for obtaining an estimate of the flow at the glottis. Apparently, the glottal flow \( U_g \) is related to the flow in the trachea \( U_t \) and in the pharynx \( U_p \) by:

\[
U_g = U_t - C_L \frac{dP_L}{dt} - U_t - U_C \\
U_g = U_p + C_p \frac{dP_p}{dt} = U_p + U_C
\]

Due to articulatory movements the distance between the glottis and the sensors as well as the cross-sectional area of the pharynx may vary somewhat. Consequently, we cannot determine \( C_L \) and \( C_p \) exactly. Nevertheless, adopting some optimality criterion, we can estimate \( U_C \) and \( U_C \), and thereby the glottal flow. If we write:

\[
C_L = (A_L - Ax)/(pq_c^2) \\
C_p = (A_p - (10-\Delta x))/(pq_c^2)
\]

where \( Ax \) is the distance between the glottis and the location in the middle of sensor #1 and #2, \( A_L \) and \( A_p \) are the cross-sectional areas of trachea and pharynx resp., and \( c \) is the sound velocity, we may rewrite eq. (3) as:

\[
U_g = - \frac{A_t}{5p_0} \left/ \left( (P_3-P_2)dt \right) \right. - 5 \Delta x/c^2 \cdot \frac{dP_2}{dt} \\
U_g = \frac{A_p}{5p_0} \left/ \left( (P_3-P_4)dt+5\left(10-\Delta x\right)/c^2 \cdot dP_3/dt \right) \right.
\]

Since we do not know the exact value of \( \Delta x \), this parameter can be used to optimize the estimates of \( U_g \). The optimality criterion we used is based on the assumption that \( U_g \) should be constant during the interval that the membranous parts of the vocal folds are pressed together over the entire glottal depth. In practice this means that \( U_g \) must have a flat portion during at least part of the closed glottis interval. Note that the estimates of \( U_g \) in eqs. (5) are not independent and, that, ideally spoken, it should be possible to find one value for \( \Delta x \) where both estimates become identical.

RESULTS

Using eqs. (5), we calculated estimates of \( U_g \) from subglottal and supraglottal pressure gradients. We did so for many different /a/ and /i/ vowel segments. Very consistently we found that for the vowel /a/ one and the same \( \Delta x \) yielded optimal estimates of \( U_g \) (in the sense of the optimality criterion specified above), and that these \( \Delta x \) values were close to 0 cm. It should be noted, however, that the longitudinal impedance could not be assumed to be purely inductive as indicated in Fig. 1 but that some viscous damping had to be added both in the trachea and the pharynx in order to obtain slopes in the closed glottis intervals of \( U_g \) which were close to zero. The best results were obtained if the ideal integrators in eqs. (5) were replaced by leaky ones with a transfer function \( k(t)=1/(1+4at) \), with 975 < \( a < 99 \).

Because the subglottal system can be assumed constant and, due to the relatively high glottal impedance, virtually uncoupled to the vocal tract, we concluded that estimates of \( U_g \) which are derived from the subglottal pressure gradient are valid estimates regardless of the vowel produced. And indeed, as we had expected, also for /u/ and /I/ vowels it appeared possible to obtain credible estimates for \( U_g \) from the subglottal gradient signals with values of \( \Delta x \) which were of the same order of magnitude as for /a/.

However, for /u/ and /I/, the estimates of \( U_g \) derived from the supraglottal gradient signals could not be made identical to the estimates derived from the subglottal gradient signals. Some experimentation with the value of \( \Delta x \) that would be needed to optimize the waveform of \( U_g \) derived from the supraglottal pressure gradient showed that any non-zero value for \( \Delta x \) (i.e. \( \Delta x = 10 \) cm) made the estimate worse. The wave caused by the fact that in practice the waveforms of \( U_g \) for /u/ and /I/ have a flat closed interval, so that "compensation" with \( dP_3/dt \) only gives rise to the introduction of a first formant ripple in \( U_g \).

Because there is no reason to doubt the validity of the supraglottal pressure recordings we are left with two possible explanations for the problems: either the electrical model of Fig. 1 is too crude and simplistic or the assumption of plane wave propagation in the vocal tract is too crude for /u/ and /I/ vowels. In order to check the possible explanations we have implemented a speech production model on a computer. In the model both the number of sections and the electrical equivalent of such a section can be varied.

The model was driven by a constant pressure source, the internal impedance of which was calculated by means of a modified two-mass model of the vocal folds. Vocal tract geometries for the vowels were taken from Fant's Russian vowel data. This model was used to calculate pressure waveforms at 2.5 and 7.5 cm above the glottis. Subsequently these waveforms were processed in the same way as the measured signals. The real flow at the input of the model was also calculated.

The estimates of the glottal flow that were obtained from the "pressure gradients" 5 cm from the source could hardly be distinguished from the real glottal flow at the input of the model. This result appeared to be true for all vowels, irrespective of the number of sections used to model the tract or the presence or absence of struts in the network that represent wall impedance. The only way to obtain flow signals with nearly zero noise during the closed glottis interval at a point 5 cm from the glottis was to make the cross-sectional area of the lowest 5 cm of the pharynx practically zero. This is obviously in discord with all X-ray data available.

On the basis of these results, we are now inclined to believe that eq. (1) does not give an appropriate description of the air behavior, at least not in the first 5 cm of the pharynx. A possible explanation would be that a jet is formed at the glottis which takes about 5 cm before it has completely expanded and has attached to the wall again, so that in fact the position where the actual excitation takes place is shifted downstream. Note that assuming an effective source located somewhere in the pharynx where above the glottis would question the all-pole character of the resulting vocal tract transfer function. It might also explain the often reported difficulty of obtaining glottal flow pulses with a flat closed glottis interval by means of inverse filtering, especially for vowels with a low F1.

ACKNOWLEDGEMENT

This research was supported by the Foundation for Linguistic Research which is funded by the Netherlands Organization for the Advancement of Pure Research, ZWO.
RELATIONSHIPS AMONG PARAMETERS OF THE GLOTTAL WAVEFORM AND INTENSITY VARIATION FOR MALE AND FEMALE SPEAKERS

E.B. Holberg1,2, R.E. Hillman1,3 and J.S. Perkell2

(1) Department of Communication Disorders, Boston University, 48 Cummington St., Boston, MA 02215, U.S.A.
(2) Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

INTRODUCTION

There are a number of studies in which glottal wave parameters have been examined for small numbers of subjects (cf. Tímoke et al., 1958, Guelffin and Sundberg, 1978). Such studies have contributed greatly to understanding the details of vocal function. However, in order to establish normative data as a basis for evaluating pathological voice function, it is also necessary to study larger groups of normal subjects. The present study examined relationships among selected glottal wave parameters and intensity variation for 16 male and 16 female normal adult speakers. This work was carried out as part of an ongoing study of normal and pathological voice production.

METHODS

All subjects were non-smokers and had no history of vocal pathology. The speech material consisted of strings of five repeated productions of the syllable /pæ/. The subjects were asked to produce the strings in a "comfortable" normal, soft, and loud voice. The strings were repeated five times for each speech condition.

Recordings were made of oral air flow and sound pressure. A high time resolution pneumotachograph attached to a circumferentially vented face mask was used to transduce oral volume velocity (flow) (Rothenberg, 1973). An external microphone at a fixed distance from the subjects' lips was used for the sound recordings. The flow and acoustic signals were recorded, along with calibration signals, on FM tape. At the time of the recording flow was low-pass filtered at 900 Hz. In order to eliminate the effects of vocal tract resonances above the first formant (F1), At A/D conversion, the flow signal was low-pass filtered again at 900 Hz. For each speech condition, the token with an SPL value closest to the mean for the 15 tokens was chosen for inverse filtering. The flow signal for the selected token was inverse filtered with a single zero pair having a center frequency at the value of F1 and a bandwidth of 70 Hz. (Pent, 1972). The resulting "glottal wave" signal was differentiated and smoothed to obtain a first derivative or "velocity" signal. For each of 4 cycles, data were extracted interactively at the time of: beginning of opening, peak opening, closure and maximum closing velocity.

From the extracted data, the following measurements were calculated: open quotient (OQU), which is the ratio of open time to time of the entire cycle, speed quotient (SQU), the ratio of the opening time to the closing time, the period (PER) and maximum velocity of closing (VCLL), measured in liters/sec^2. These measures were averaged over the 4 cycles.

Intensity data were extracted interactively from a smoothed version of the mean square of the sound pressure signal.

RESULTS AND DISCUSSION

Relationships between the glottal wave parameters and SPL were examined in two ways: 1) within each intensity condition, and 2) across intensity conditions as a function of increasing intensity from normal to loud voice and decreasing intensity from normal to soft voice. Statistical treatment of the data included simple (pairwise) and multiple linear regression analyses of the male and female groups separately. Analysis of variance of multiple regression coefficients over the subject groups (male vs. female) was also performed in order to evaluate male-female differences.

Within each speech condition

Mean intensity value and standard deviation (in parentheses) for normal intensity were 78.8 dB (6.2) for male speakers and 77.3 dB (4.4) for female speakers. For loud voice, means and standard deviations were 85.7 dB (4.2) for males and 84.1 dB (5.1) for females, and for soft voice, 72.2 dB (3.6) for males and 71.7 dB (4.5) for females. For most male and female speakers, fundamental frequency increased in loud voice and decreased in soft voice.

Analysis of variance of multiple regression coefficients over the groups did not show significant differences between males and females within any of the intensity conditions. However, some qualitative observations could be made: Mean values for VCLL and OQW were higher, and values for OQW were lower for males than for females in all speech conditions. For both males and females, mean values for VCLL were highest in loud voice and lowest in soft voice, and mean OQW was lowest in loud voice and highest in soft voice. In spite of the fact that the standard deviations for these parameters were large, these results generally support the view that glottal waveforms for male voices have a higher VCLL and relatively larger closed phase than glottal waveforms for female voices (cf. Mønse and Rygebritsen, 1977).

Changes in SPL and glottal wave parameters

A summary of the direction of change in glottal wave parameter as a function of increasing and decreasing intensity is shown in Table 1. The numbers in the table correspond to number of subjects. When the number of subjects do not add up to 16, there was no change in the parameter for the remaining subjects.

<table>
<thead>
<tr>
<th>From Normal to Loud</th>
<th>From Normal to Soft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>PER</td>
<td>2</td>
</tr>
<tr>
<td>OQUW</td>
<td>4</td>
</tr>
<tr>
<td>SQU</td>
<td>7</td>
</tr>
<tr>
<td>VCLL</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 1. Change in parameters for normal to loud and normal to soft voice for 16 male and 16 female speakers. + indicates increase and - decrease of parameter value.
Increasing intensity from normal to loud voice

Similarities between males and females: Most male and female speakers increased fundamental frequency with intensity. For 10 of 16 males, VCL was higher in loud voice as compared to normal intensity. The simple correlation (r) between change in SPL and change in VCL was .75. For females, VCL increased for 14 of 16 speakers, and a squared .71. SPQV did not vary systematically with SPL for either male (r=.19) or female speakers (r=-.10).

Differences between males and females: There was a clear difference between males and females in the parameter OPQV, which decreased for 12 of 16 males (r=-.68), but increased for half of the female speakers and decreased for the other half (r=.08). The multiple correlation between change in SPL and change in the combined set of glottal wave parameters for increasing intensity was .89 (p<.001) for males and .81 (p=.01) for females. For males, increase in SPL was primarily related to decreased OPQV (p=.01), increased SPQV (p=.07), and to a lesser degree, increased VCL (p=.12). Conversely, for females, the only parameter (except for period) which was systematically related to increased SPL was increased VCL (p=.01).

The difference between males and females in terms of relationships between changes in the set of glottal wave parameters and increased intensity was verified by an analysis of variance of regression coefficients (p=.07). Further experimental and theoretical work is needed to determine whether or not this observed difference is related to structural and/or functional differences between the vocal mechanisms of males and females.

Decreasing intensity from normal to soft voice

There did not appear to be any noticeable differences between males and females for decreased intensity, either qualitatively or in terms of the simple correlations between change in SPL and change in each of the individual glottal wave parameters. 10 of 16 of the male and 12 of 16 female speakers lowered the fundamental frequency in soft voice. For all male speakers VCL was lower for soft voice than for normal voice (r=.54). For 11 of 16 female speakers VCL decreased in soft voice (r=.77). For 15 of 16 males OPQV increased in soft voice. However, the simple correlation between change in OPQV and change in SPL was -.11. This low correlation indicates that, although most male speakers showed an increase in OPQV for soft voice, the relationship between OPQV and decrease in SPL was not linear. OPQV decreased for 11 of 16 females (r=.41). SPQV decreased for 14 of 16 males (r=.49) and for 12 of 16 females (r=.41).

The multiple correlation between change in SPL and changes in the combined set of glottal wave parameters for decreasing intensity was .71 (p=.06) for males and .85 (p=.001) for females. In this analysis, for males, only decreased VCL (p=.11) showed marginal evidence of being systematically related to decreased SPL. For females, decreased SPL was primarily related to decreased VCL (p=.003) and increased OPQV (p=.09).

Even though the glottal wave parameters of VCL and OPQV appeared more highly related to decreased intensity for females than for males, analysis of variance of regression coefficients over the groups did not show a significant difference between males and females.

We speculate that additional factors such as air flow may have an influence in the production of soft voice which is different from normal or loud voice. An increase in flow and a lower AC/DC value would result in a more breathy and softer voice with a steeper spectral slope.

SUMMARY

The present study examined relationships among selected glottal waveform parameters and intensity variation for normal male and female adult speakers. Differences between males and females were found for change in intensity from normal to loud voice. The results of multiple regression analysis showed that for male voices, increased SPL was primarily related to decreased open quotient, increased speed quotient and, to a lesser degree, increased velocity of closing. For female voices, the only parameter that was systematically related to increased SPL was increased velocity of closing. Although velocity of closing and open quotient were related to decreased intensity to different degrees for males and females, no statistical significant difference between males and females was found for decreased SPL.

ACKNOWLEDGEMENT

This research was supported by NIH Grant No. NS21183.

REFERENCES


EFFECTS OF FREQUENCY SCALE TRANSFORMATION OF SPEECH SPECTRUM

Kazu Nakata and Kho-Dish Nishino

Tokyo University of Agriculture and Technology,
2-24-16, Nakamachi, Koganei-shi, Tokyo, JAPAN

INTRODUCTION

Recent developments of speech technology mainly deal with a phonetic/phonemic aspect of speech, and a quality aspect of speech is left almost un-revealed. And all of the difficulties now we are confronting, such as speaker independent speech recognition, natural sounding speech synthesis from text, and low bit-rate speech coding with high quality, are deeply related to the quality of speech.

We first define a quality of speech and discuss a way of its research, and a new high quality speech synthesis method is introduced for the purpose.

By the method, frequency scale of spectrum, duration and pitch of an utterance are independently transformed linearly without noticeable distortion or noise due to the transformation.

As our first approach to the quality problems of speech, some experimental results of frequency scale transformation are presented and discussed.

A QUALITY OF SPEECH

Generally speaking, speech has two aspects, i.e., phonetic/phonemic or linguistic aspect and quality aspect.

Phonemic aspect is measured or evaluated by an articulation or intelligibility tests. These concept and technique have been developed in the history of telephone transmission of speech, where only a human speech voice was concerned. This synthesis naturalness of speech or human-like sounding nature of speech is always assmed.

Under those circumstances, a quality of speech simply means conservation of speaker identity.

But recent developments of speech technology often replace a human speech voice by a synthetic speech. Then a quality of speech means a human-like sounding nature in the first and speaker identity in the second.

In the case of a quality evaluation of synthetic speech, especially synthetic speech from text, the human-like sounding nature is an essential prerequisite of its practical applications.

Then the establishment of a new concept of speech quality which is applicable not only to natural speech but also to synthetic speech is strongly required for an establishment of future speech technology application fields.

Human-like sounding nature and speaker identity are mutually related deeply and a quality aspect of speech plays an essential role in the technical realization of the true speaker independent speech recognition.

THE WAY OF RESEARCH

Phonetic/phonemic aspect of speech and quality aspect of speech is also deeply related, but in the first approximation, we can assume they are independent.

Physical correlates of the phonetic/phonemic aspect of speech are its spectrum and some voice source characteristics, but mostly spectrum.

The statement is supported by the fact that almost all of the present speech recognition techniques and devices use only spectrum features of speech waveforms and non of them use voice source features positivly.

For the quality aspect of speech, voice source characteristics play important roles and the fact is supported by the developments of many techniques which utilize a residual source waveform to improve voice quality of LPC analysis-synthesis system.

A speech analysis technique of present level can not separate these two factors from each other, and analyzed data related to both aspects of speech [1].

Normally, we analyze speech waveforms and try to find out physical correlates of speech quality among them. But physical correlates of speech quality are not so simple that they can be clearly revealed by the present level of analyzing technique.

On the other hand, synthetic quality of speech becomes so high that they can be used in the finding out of physical correlates of voice quality by the Independent control of some physical feature of speech waveforms and listening evaluation of its effects on speech quality.

A NEW METHOD OF SPEECH SYNTHESIS

Ordinary LPC analysis is carried out in the way that a certain number of speech samples are processed in one block, and a block length usually ranges from 10 to 30 ms. A control of spectrum, pitch or duration must be done in the unit of this block length, and this becomes a main reason of quality degradation of the transformed speech.

Recently, a new analysis method of LPC based on the GIVENS transformation is proposed [2]. Data such as LPC parameters, voice source (residual) waveforms are efficiently extracted from a real speech waveforms sample by sample. And a fine control of a transformation becomes possible and a good quality of transformed speech is obtained.

We basically use this analysis method and LPC synthesis in this kind of parameter transformation of speech. Fig.1. shows the typical processes in the linear transformation of frequency scale of speech spectrum.

Firstly, speech waveforms are analyzed by LPC method based on GIVENS transformation, and data or LPC parameters and samples of residual waveforms are obtained at every sampling time. These parameters are inserted or eliminated at pre-set rate for the given transformation conditions.

An insertion of data means repeating a given small portion of data as shown in Fig.1., as well as in the case of elimination. After this date rearrangement, a synthesis by LPC lattice filters has been performed with variable clock rate.

Then duration and pitch are kept as same as the original speech but its spectrum is transformed linearly in frequency scale.

A pitch or a duration is also linearly transformed in each way without changing other parameters. A some careful adjustment is necessary to get good quality in transformed speech especially on the insertion or elimination of residual waveforms.

A certain form of smoothing with a threshold for its amplitude either the smoothing is done or not, is very effective for a quality improvement.
RESULTS OF A SPECTRUM TRANSFORMATION

Our first approach to the quality aspect of speech starts with a linear transformation of frequency scale of spectrum.

Two physiological facts in speech waveform production are vocal tract resonance and voice source generation. They are represented physically by speech spectrum envelope and its fine structure as shown often in their FFT spectrum.

The first experiment consists of a quality evaluation of a linearly spectrum transformed speech. These samples are transformed only the frequency scale of spectrum, and all other features of speech are kept same as possible.

Details of the experiment are described in the followings.

EXPERIMENT-#1: Differential limen test for voice quality.

Experimental conditions:
1) test speech: one short Japanese sentence of one male speaker (about 3 sec in length).
2) transformation: a linear transformation of frequency scale of spectrum and all other factors are kept as original one as possible.
3) range of transformation: ±5% with 1% step.
4) test method: A-B pair comparison, and always either one of A or B is the original one.
5) instruction for hearing tests:
   Q: B is the same voice of A, and A: yes or no.
6) listeners: ten adults (one female) who have normal hearing and a half of them know the speaker and the others are not.

Experimental results: shown in the Fig. 2.

EXPERIMENT-#2: Differential limen test for a personal identity.

Experimental conditions:
1) test speech: the same as #1.
2) transformation: the same as #1.
3) range of transformation: ±13% with 2% step.
4) test method: the same as #1.
5) instruction for hearing tests:
   Q: B is the voice of the same person of voice of A, A: yes or no.
6) listeners: the as #1.

Experimental results: shown in the Fig. 3.

DISCUSSIONS ON THE EXPERIMENTAL RESULTS

Fig. 3 and 4 tell us that the DL of a linear frequency scale transformation regarding to voice quality is about ±2% and the DL for individual identity is about ±5%. These results are in good coincidence with a same kind of experiments carried out at NHK Technical Laboratory by Kuwabara and Ohgushi [3].

The only difference between these two experiments are test sentence and method of frequency scale transformation.

We are now also carrying listening tests of frequency scale transformation of spectrum with a voice source transformed in the same manner as the spectrum and the duration is kept as same as possible of the original one.

Results will be presented at the conference.

Acknowledgement

The authors express their appreciations for the financial support given by HOSO-BUNKA foundation.

References
GENERATION OF CONTROLLED SPEECH STIMULI BY PITCH-SYNCHRONOUS LPC ANALYSIS OF NATURAL UTTERANCES

M. J. Hunt and C. E. Harvenberg
National Research Council of Canada, National Aeronautical Establishment, Montreal Road, Ottawa, K1A OR6, Canada

1. INTRODUCTION

The purpose of this paper is to point out the potential of pitch-synchronous linear predictive coding (LPC) for the generation of controlled stimuli for speech perception experiments and to provide practical information concerning the implementation of pitch-synchronous LPC techniques. Much of what is said applies equally to the use of pitch-synchronous techniques for formant analysis and in speech output systems generating announcements.

Stimuli are frequently generated by formant-based synthesizers. The parameters of the synthesizer — fundamental frequency, formant frequencies and bandwidths, and so on — are modified in various ways and the effect on perceptual judgments is studied. The starting values of these parameters are derived either from a set of rules or from careful hand copying of spectrograms augmented by many cycles of listening to the result and readjusting parameters. The latter method can give results that are essentially indistinguishable from the original, but it can also be an extremely lengthy process. The former, in which formant trajectories are computed by a formula, frequently produces results that do not sound natural. The extension to natural speech of conclusions drawn from experiments using such stimuli is sometimes questionable. Moreover, such methods are certainly unsuited to the study of subtler aspects of speech perception involving, for example, judgments of a speaker's identity or emotional state.

The kind of analysis carried out in LPC vocoders [1] is in principle capable of deriving the parameters automatically from a natural utterance. Unfortunately, the need for real-time performance and for robustness to degradations in the speech being analysed impose compromises on the algorithms used in vocoders that result in an unreliable formant analysis and in a synthesized speech that does not sound natural.

There have been a number of developments that improve naturalness at the expense of bit rate by providing more detailed excitation information, a recent and popular example being multipulse LPC [2]. These developments do not improve the reliability of formant information and so introduce into the glottal cycle coded differently, they may, in some cases, lead to voiceless speech due to noise. However, we have found that in certain /h/ sounds that are phonetically breathy vowels there is no detectable voiceless component and there is a negative peak. Synthesis is better if these sounds are treated as voiceless, so a preliminary decision to treat a period as voiceless is reversed if a negative peak is detected. Also, as voiced regions tend to have a lower voice level, sometimes, what appears to be voiceless is actually voiced in the multipulse models.

3. PITCH-SYNCHRONOUS LPC

In so far as the acoustic effect of the vocal tract can be modeled as an all-pole filter, covariance method LPC analysis carried over suitable portions of voiced speech related to the instants of main excitation on glottal closure can in principle be used to derive the parameters of the filter. In practice, vocoders carry out their analyses over regularly spaced portions of the signal spanning two or more glottal cycles, and the covariance method of analysis, which can in certain circumstances give rise to unstable filter parameters, is replaced by the less exact but more robust autocorrelation method. In addition, vocoders are subject to errors in the voicing decision and in the estimation of fundamental frequency.

In the schemes we describe here each of these deviations from the ideal is avoided. Recordings are made in quiet non-reverberant conditions. Synchrony with glottal excitation is achieved by simultaneous recording of the output from a laryngograph [3], a device which measures the r.f. impedance across the larynx and hence indicates the degree of contact of the vocal cords. Covariance method LPC is used to estimate a suitable transfer function on a computer. Any resulting instabilities are detected and corrected.

Recordings are made in an anechoic chamber using a condenser microphone and a p.m. coder unit generating a signal suitable for recording on a video recorder. This provides low-noise recordings free of amplitude and phase distortion. The laryngograph signal is recorded on the second channel.

In order to transfer these recordings to our computer, they have to be digitized. This is carried out at a 20 kHz sampling rate with a two-channel 12-bit a/d converter. Before digitization the signals are passed through low-pass filters with a cutoff frequency set at 7.5 kHz to ensure negligible aliasing in the negligible phase distortion in the frequency region up to 5 kHz.

The redigitized recordings are low-pass filtered with a tenth-order Butterworth filter with its 1.25 dB point at 5 kHz. This process is carried out twice, once forward recording the signal twice, and once backwards, which is equivalent to using the forward recording with its 3 dB point at 5 kHz, but with no phase distortion. The recordings are then subsampled to 10 kHz.

The laryngograph signal is time-differenced. Glottal closure then appears as a positive spike, and the start of reopening as a much smaller and blunter negative peak.

Because of the time taken for an acoustic signal to propagate from the glottis to the microphone (of the order of a millisecond), there is an offset between the speech and laryngograph signals. This offset can be determined and removed by carrying out a preliminary LPC analysis of small portion of voiced speech and aligning the LPC residual with the laryngograph spike, since they both correspond to the instant of glottal closure.

The zeroth-order autocorrelation coefficient of the laryngograph signal is the primary indication of voicing and fundamental frequency. If no such spike is found in a period corresponding to the longest possible glottal period (i.e., about 20 ms), the period is tentatively classified as voiceless. However, we have found that in certain /h/ sounds that are phonetically breathy vowels there is no detectable voiceless spike but there is a negative peak. Synthesis is better if these sounds are treated as voiceless, so a preliminary decision to treat a period as voiceless is reversed if a negative peak is detected. Also, as voiced regions tend to have a lower voice level, sometimes, what appears to be voiceless is actually voiced in the multipulse models.

Before being analysed, voiced speech is pre-emphasized by sample-by-sample time differencing with a 0.95 leak factor. Tenth-order covariance method analysis is then carried out on portions of the speech signal determined by the aligned laryngograph signal. In unvoiced speech, we fix the LPC analysis window.

With voiced speech, we obtain best results when the analysis window starts ten (i.e., order) samples before the instant of closure and ends just before the next instant of closure. One might have expected that the best results would be obtained with an analysis confined to the closed-glottis phase, and Krishnamurthy [5] has indeed reported finding this to be the case for formant analysis purposes. In our tests, we have found a clear advantage for formant analysis in using just the closed-glottis phase, but we have, on the other hand, found a clear advantage in using the full glottal cycle as far as the detailed representation of the synthesized speech is concerned. We suspected that the poor results with the closed-glottis analysis might be due to an underestimation of effective formant bandwidths (since bandwidths increase as frequencies fall slightly during the open phase), but increasing the bandwidths did not help. Another possibility was that the shorter analysis window did not make the noise visible in the noise; but, again, correcting for this by combining autocorrelation data over consecutive glottal cycles did not help.

To check for instabilities in the filter derived by the covariance analysis, we solve the polynomial whose coefficients are the predictor coefficients. For this purpose we use the fortran subroutine polr from the widely distributed Scientific Subroutine Package with some checks added to prevent floating point overflow and
underflow. The imaginary and real parts of the roots, when converted from $z$-plane to $s$-plane coordinates, are the formant frequencies and bandwidths. Instabilities manifest themselves as negative bandwidths, and they can be corrected simply by reversing their sign. At this point, the formant frequencies and bandwidths can be modified before being reconverted into predictor coefficients for synthesis.

In the resynthesis the predictor coefficients are updated every glottal cycle. Unless a voicing boundary is being crossed, the ten-sample memory in the filter is loaded with the last ten samples of the output from the previous cycle, and a suitably scaled version of the excitation function is then passed through the recursive filter. Finally, voiced periods are de-emphasized by leaky integration.

The excitation function can be scaled so that the power in the synthesized speech matches the power in the corresponding portion of the input speech. The scale factor used can be calculated by techniques of the following manner. The current cycle is first synthesized with the amplitude of the excitation function set to a standard value and the filter memory initialized to zero. This will produce a vector of samples $a$. Next, it is synthesized with the filter memory loaded with the last ten samples of the previous cycle synthesized with correct power and the excitation amplitude for the current cycle is set to zero, producing a second vector of samples $b$. The final synthesized waveform will be of the form $ga + b$, where $g$ is the scale factor on the excitation. Writing the corresponding original speech waveform as $s$ and setting the powers equal

$$\langle s^2 \rangle = \langle g^2a^2 \rangle + \langle b^2 \rangle + 2g\langle ab \rangle$$

This can be solved for $g$. We have found that it is best to carry out the power estimation over the first 75% of a glottal cycle, since, particularly at the onset of voicing, there can be excitation activity before the next instant of closure that is permanent associated with the next cycle and that would cause the power in the current cycle to be overestimated if it were included.

The excitation function for unvoiced speech is Gaussian noise, and that for voiced speech can be a simple impulse. We have found, however, that while this gave speech of high intelligibility, the voiced speech seemed to be lacking in very low frequencies. The replacement of the impulse with a double-differentiated version of a function modeling glottal airflow (waveform $c$ in the paper by Rosenberg [8]) helped to some extent, but the best results have been obtained with a waveform produced by aligning and averaging together the LPC residual from several glottal cycles of one or more vowels.

3. PROSPECTS FOR SIMPLIFYING THE METHOD

It is possible that the instants of glottal closure could be determined from the residual of a preliminary pitch synchronous LPC analysis, thus eliminating the need for the laryngograph, but such a method seems unlikely to be reliable.

To test the need for the anechoic chamber, we processed a recording digitally to simulate the reverberation introduced by a typical small room and then carried out a pitch-synchronous analysis and synthesis. The resynthesized speech was of much worse quality than that from the reverberance-free counterpart.

To test the need for digitization at 20 kHz and for careful filtering, we simulated the effect of a six-order low-pass filter set at 4.5 kHz, such as might be used before digitization at 10 kHz. The synthesis obtained from a recording processed in this way was noticeably worse than that of its counterpart sharply filtered at 5 kHz, and we were able to conclude that 20 kHz sampling, or at least very sharp analogue filtering at 5 kHz, is desirable. On the other hand, the need for low-pass filtering before and after digitization to eliminate phase distortion is much less apparent: filtering the recording twice forwards instead resulted in synthesis whose difference from the its counterpart filtered forwards and backwards was barely detectable.

To test the need for a digital tape recorder, we simulated the effect of the low-frequency phase distortion introduced by direct-recording analogue recorders [7]. We found it had no effect at all on the synthesis. On the other hand, the noise introduced by the use of an f.m. analogue recorder had a large effect on the synthesis. We suspect that pitch-synchronous analysis is extremely sensitive to noise and that our method would be improved if our 12-bit a/d converter were replaced by a 14 or 16-bit version.

4. EXPERIENCE IN GENERATING STIMULI

All voiced sentences without nasals synthesized by the method just described sound very natural at least on casual listening and can be taken for the original recording, though when under careful comparison differences are apparent. In more difficult sentences, faults are audible, but the synthesis is still in our judgment much better than conventional LPC or synthesis-by-rule. The formant and excitation parameters contain very few errors and can be used as a quickly obtained starting point for a manually or automatically manipulated formant parameters, for example, and then re-excite with the LPC residual. This provides modified speech with no audible faults.

Timing and fundamental frequency changes can be carried out easily. Exciting the voiced portions with time-differeenced noise provides an entirely convincing whisiper. A reasonably convincing female can be produced from a male original by modifying the excitation and scaling all the formant frequencies; while modifying individual formant frequencies can produce an apparent dialect change. In an extreme modification the timing of "We were away a year ago" to match the first few bars of Beethoven's Fifth Symphony played by the Berlin Philharmonic Orchestra and used the music as the excitation signal. The result was a perfectly intelligible sentence apparently spoken by the instruments of the orchestra.

In work being carried out with R. Mowrey and T. Planton at the Ottawa General Hospital, we have produced stimuli with formant values interpolated between /ba/, /da/ and /ga/. The results continue to sound natural and the speaker characteristics are retained, even though the sounds produced are physically impossible.

These stimuli contain voiced murmurs occurring before the release of the stops. Since the vocal tract is completely obstructed, LPC analysis would not be expected to work well. Surprisingly, however, if additional pre-emphasis is applied to flatten the spectrum in the pre-release regions, synthesis could be obtained that sounded like the original, produced spectrograms that looked like the original, and had plausible, continuous formant tracks. It seems possible that analysis of nasals may also be improved with modified pre-emphasis.

ACKNOWLEDGMENT

This work was partly funded by the Department of National Defence, Canada.

REFERENCES

CONTROL OF SPEECH SYNTHESIZER PARAMETERS

R D Wright, S J Elliott

Institute of Sound and Vibration Research, The University, Southampton SO9 7JR

INTRODUCTION

A speech synthesizer is a set of parameters representing some model of speech production or speech acoustics. A particular set of parameter values determines a particular sound quality. A sequence of sets of these target values can then be used to produce continuous speech.

Attention has been given to good target values, and to modification of target values as a function of phonetic context (coarticulation). This paper concentrates on what happens during the transitions between targets.

When nearly all synthesizers were formant synthesizers the transition or interpolation problem was basically one of approximating the formant motion observed in speech spectrograms. With the advent of various linear prediction parameters and related representations, the situation is not so straightforward. Given that we know what happens when we linearly interpolate resonance parameters, what happens to resonances when we linearly interpolate other parameter types?

Interpolation Methods

It was originally intended to study two aspects of parameter control: parameter type and interpolation method. Various interpolation methods have been used, and proposed for speech synthesis: piecewise linear (Klatt, 1979; Holmes, et al, 1969); decaying exponential (Rabiner, 1969); increasing exponential (Lawrence, 1972). Other synthesis has used linear interpolation on a log frequency scale, which is either increasing or decreasing, depending upon the direction of the transition.

These methods appear to be quite different. However all such methods produce the same path in parameter space, as shown in Figure 1 for linear and exponential interpolation of F1 and F2. It can be seen that the paths in the F1 vs F2 space are identical; it is simply the time sampling (or rate of motion) along the path that differs. In this study only linear interpolation was used.

**Figure 1**

A) Linear and exponential interpolation of formant parameters as functions of time.  

B) Same data plotted in F1 vs F2 space. Transitions follow same path.

METHOD

The starting point was series resonance data taken from Klatt (1980). This gave center frequency and bandwidth targets for the phonemes in eight nonsense words. Then for each synthesizer a conversion was performed to determine targets in that synthesizer's own parameter type. Linear interpolation was then used to determine intermediate values between target points. Finally, results were converted back to formant frequencies and bandwidths and plotted as stylised spectrograms.

Parameter Conversion

Four of the six synthesizers studied are all-pole models for which an analytic conversion relationship exists; these are:

- series resonance;
- direct form (prediction coefficients);
- reflection coefficients;
- area function;

The two remaining types (parallel resonance and articulatory parameters) involve approximations. An exact conversion from series to parallel form is possible through partial fractions, but only for one particular set of amplitudes and bandwidths. If bandwidths are fixed and amplitudes are variable then a parallel synthesizer can be made approximately equivalent to a series form by using the amplitudes at resonance of the series form.

The conversion from resonance data to articulatory parameters followed the method of Ladefoged et al. (1978). This uses regression equations to estimate tongue and lip parameters from three formant frequencies. Alternatively, the Harrisman et al. (1977) X-ray data basis vectors can be used to convert from area function to tongue parameters.

Speech Sound Categories

To cover the possible combinations in an efficient way the following strategy was used:

1. Six categories were considered: vowel, approximant, fricative, stop, nasal, and affricate.
2. Only extreme values in each category were used. Thus the 18 vowels and diphthongs were represented by /i/, /a/, and /u/.
3. For the stop, fricative and nasal groups, extreme formant values were mainly associated with front place of articulation (and with voicing in the case of stops and fricatives); thus only /v/, /b/, and /m/ were selected, plus /j/ to reach the extreme of the second formant range for stops. Affricates were discarded as they have the same transitions as the stop/fricative components from which they are synthesized.

4. Each chosen consonant need not pair with each of the three vowels: two pairs span the extremes in most cases.

5. Finally, three transitions can be nicely represented on one plot, so sequences of up to four sounds can be analyzed at once, reducing the number of plots produced.

This strategy resulted in the following nonsense words: /lau/, /dju/, /waju/, /rilu/, /viva/, /iba/, /agu/ and /ima/.
RESULTS

A typical result is shown in Figure 2, which compares a series resonance synthesizer with synthesis using an area function representation. Note that both synthesizers reach identical formant values at the target points, as an exact conversion is possible between center frequency and bandwidth data (the series resonance parameters) and the area function data. The transition paths between targets are far from identical, however.

General results for all the tokens and all the synthesizers are summarised as follows:

<table>
<thead>
<tr>
<th>Synthesiser</th>
<th>Transition Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series</td>
<td>(Reference path)</td>
</tr>
<tr>
<td>Parallel</td>
<td>Amplitude differences</td>
</tr>
<tr>
<td>Direct form</td>
<td>Instability</td>
</tr>
<tr>
<td>Reflection coeffs</td>
<td>Bandwidth differences</td>
</tr>
<tr>
<td>Area function</td>
<td>Frequency and bandwidth differences</td>
</tr>
<tr>
<td>Articulatory</td>
<td>Large bandwidth differences</td>
</tr>
</tbody>
</table>

The most marked effect of interpolation was observed for the direct form during two of the eight nonsense words: the resonance damping became negative, corresponding to instability in the steady state. This shows that potential instability can arise simply by trying to get from one stable position to another; this problem does not affect reflection coefficients and area functions. A stability problem can also arise with tongue parameters, but only if negative vocal tract areas are allowed.

A less disastrous but more general effect is the observation that bandwidths are proportionally more affected than frequencies.

In the original series resonance data, the upper two formants (F4 and F5) did not vary. These fixed resonances remain fixed for interpolation in the direct form. They began to move slightly when reflection coefficients were used, and were very much affected when the representation was in terms of area functions.

Finally, the method used to produce articulatory parameters was based only on formant frequencies, not formant amplitudes or bandwidths. It is not surprising, therefore, that the resultant vocal tracts had formant bandwidths which were very variable, and not a good match to the Klatt data even at the target points. Thus large overall bandwidth differences were observed.

Figure 2: Results for two synthesizers, for the word /rulu/. The data are the formant paths resulting from linear interpolation of 1) series resonance parameters; 2) area function.

DISCUSSION

A direct comparison of synthesizer parameters and their formant transitions reveals certain effects as summarised above. A more general critique, however, has broader considerations: simplicity, independence and physical interpretation of control parameters. When viewed in these terms, there are problems with a formant representation:

1. Formant parameters are used in synthesis as independent control variables, whereas physical formants are not independent. For instance, a change in tract length affects all the resonances simultaneously. Thus formants are not a minima dimensionality set of control parameters.

2. Similarly, a simple change in vocal tract shape produces a complicated set of changes in the formant motion. This is well known. For example, Fant (1960) shows how a simple motion of a constriction from one end of a tube to the other produces a sequence of changes in five formants.

3. Finally, formants cannot be made subject to dynamic constraints in any simple fashion. Formants don’t move, it is the tract which moves. Simple constraints concerning articulator position and velocity are related in a highly non-linear way to formant positions.

The same criticism can be made for all the synthesizers studied, with the exception of the articulatory parameters. Traditionally articulatory models have not been widely used, partly owing to their usual complexity. The simple model evaluated in this study has no such problems, as it is merely a low-dimensionality approximation to the series resonance model or any of its exact all-pole equivalents. With digitally-implemented synthesizers it is possible to mix an articulatory and acoustic description to use tongue and lip parameters for a compact description subject to dynamic constraints, and then convert to formants for final synthesis. This approach will be the subject of further study.

Grateful acknowledgement for support is given to the SERC and the IBM (UK) Scientific Centre.

REFERENCES


Klatt D H. Software for a cascade/parallel formant synthesizer. JASA 67(3), 971-95, 1980

Ladefoged, P., Harshman, R., Goldstein, L and Rice, L. Generating vocal tract shapes from formant frequencies. JASA 64(4), 1027-1035, 1978

Lawrence, W. The phoneme, the syllable and the parameter track. Proc Speech Comm Seminar, Stockholm, Aug 1-3, 1974

ARTICULATORY SYNTHESIZER FOR RESEARCH IN LOW BIT-RATE CODING OF SPEECH

Juergen Schroeder and Man Mohan Sondhi

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

1. INTRODUCTION

Low bit rate coding of speech is an important objective of current speech research [1]. One of the most economical descriptions of the speech waveform is in terms of articulatory parameters -- parameters that specify the geometrical properties of the vocal and nasal tracts, the mechanical properties of the walls and the properties of the vocal cord oscillator. Articulatory speech synthesizers generate speech from such a description.

In the synthesizer which we have implemented, the wave propagation in the tract is assumed to be planar and linear. During the voiced portions of speech, however, the excitation of the tract is provided by a nonlinear model of the vocal cord oscillator [2] which is controlled by lung pressure, glottal rest area, intraglottal damping, pitch, and supraglottal pressure. The unvoiced portions of speech, both aspiration and frication, are generated automatically by introducing noise sources at the glottis and downstream from the narrowest constriction of the vocal tract, respectively.

At the time of this writing we have come across a recent article [3] which reports on generation of steady vowels by a method somewhat similar to ours. However, their approach does not include a model for self-oscillation of the vocal cords, ignores friction and nasality, and does not deal with the dynamic variations in the glottal parameters or in the shape of the vocal tract. All these effects are included in our model.

2. MODEL OF THE GLOTTAL SOURCE

Of the many vocal cord models that have been proposed [e.g., 2,4-7], we have selected the two-mass model of Ishizaka and Flanagan [2] since it has been shown to have very realistic properties [2,4,8,9]. The glottal volume velocity \( u_g(t) \) satisfies the differential equation

\[
R_u u_g + L_u \frac{du_g}{dt} = p_S - p_1 - p_{ng}
\]  

(1)

Here \( p_S \) is the subglottal (i.e., the lung) pressure, \( p_1 \) is the pressure downstream from the glottal expansion (see Fig. 1), and \( p_{ng} \) is a series noise pressure source located at the interface between the expansion and the first (variable) section of the vocal tract. The resistance \( R_u \) and the inductance \( L_u \) depend on \( u_g \) and upon the glottal opening. Details are given in [2] and will not be discussed here. Let us point out, however, that we solve Eq. (1) by a method quite different from that used in [2]. The most important difference is the way in which \( p_1 \) is computed. In [2] \( p_1 \) is obtained from a large system of linear equations (coupled to Eq. (1) and the oscillator equations). Instead we compute the present value of \( p_1 \) in terms of the present value of \( u_g \) and the past values of \( p_1 \) and \( u_g \). The relationship is derived in the next section.

3. MODEL OF THE VOCAL AND NASAL TRACTS

3.1 Frequency Domain Analysis

In Fig. 1, the vocal and nasal tracts are outlined. The position of the narrowest constriction between velum and lips will be needed, if the constriction is small enough, to introduce friction noise.

3.1.1 Vocal Tract. The different portions of the tract are described by four chain matrices, namely \( K_G \) for the region between glottis and velum, \( K_N \) for the nasal tract, \( K_C \) from the velum to the constriction, and \( K_L \) from the constriction to the lips. A general chain matrix of a portion of the tract relates output pressure \( p_{out} \) and volume velocity \( u_{out} \) to the input variables \( p_{in} \) and \( u_{in} \). (Upper case letters denote variables in the frequency domain.) Thus

\[
\begin{bmatrix} p_{out} \\ u_{out} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} p_{in} \\ u_{in} \end{bmatrix} = K \begin{bmatrix} p_{in} \\ u_{in} \end{bmatrix}
\]

(2)

where the input is on the glottal side, the output side is towards the lips or nostrils. For a homogeneous cylindrical tube of length \( \Delta l \) and area \( \text{area} \), the chain matrix parameters are given by:

\[
A = \cosh(s\Delta l/c) \quad ; \quad B = \frac{pc - s\sinh(s\Delta l/c)}{\text{area}}
\]

(2a)

\[
C = \frac{s}{\text{area}} \quad ; \quad D = \cosh(s\Delta l/c)
\]

(2b)

This selection of losses and wall parameters is as derived by Sondhi [10]. However, the last equation differs from Eq. (29) in [10], because we include wall compliance in the model. Each of the four matrices \( K_G, K_N, K_C, \) and \( K_L \) is obtained by concatenating elementary matrices of the type given by Eq. (2) using the same \( \Delta l \) for all elements.

At the velum two special coupling matrices are needed. For computing the chain matrix from glottis to lips the nasal side branch is represented by the matrix \( K_{CN} \) given by

\[
K_{CN} = \begin{bmatrix} 1 & 0 \\ -1/Z_{VN} & 1 \end{bmatrix}
\]

(3)

where \( Z_{VN} \) is the input impedance of the nasal branch at the velum. Similarly, for computing the chain matrix from glottis to the nostrils the vocal tract is represented by the coupling matrix \( K_{CT} \) which is the same as \( K_{CN} \) with \( Z_{VN} \) replaced by the input impedance of the oral cavity, \( Z_{VT} \).
3.1.2 Nasal Tract. The nasal tract was modelled after geometrical data of Maeda [11]. It has a geometrical length of 11 cm and a Helmholtz resonator to represent the sinus cavities.

3.2 Time Domain Synthesis

Let us now consider how the glottal model and the vocal/nasal tract model are interfaced to each other.

3.2.1 Voiced Sounds. From the previously defined chain matrices \( K_{G}, K_{N}, K_{C}, \) and \( K_{L} \) (see Fig. 1), various global matrices are formed. For example, the chain matrix from the glottis to the lips is

\[
K_{\text{tract}} = K_{L} K_{C} K_{CN} K_{G},
\]

and the chain matrix from the glottis to the nostrils is

\[
K_{\text{nasal}} = K_{N} K_{CT} K_{G}.
\]

In terms of the elements of \( K_{\text{tract}} \) it is straightforward to obtain the input impedance \( Z_{in} \) at the glottis and the transfer function, \( H_{s} \), from glottal volume velocity, \( U_{p} \), to the pressure radiated at the lips, \( p_{L} \). Similarly, from \( K_{\text{nasal}} \) one obtains the transfer function, \( H_{N} \), from \( U_{p} \) to the pressure radiated at the nostrils, \( p_{N} \). Let the inverse Fourier transforms of \( Z_{in}, H_{s}, \) and \( H_{N} \) be \( z_{in}, h_{s}, \) and \( h_{N} \), respectively. These functions enable us to express the supraglottal pressure \( p_{1} \), the pressure radiated at the lips, \( p_{L} \) and the pressure radiated at the nostrils, \( p_{N} \) in terms of \( u_{s} \). Note, that \( u_{s} \) is obtained from the nonlinear two-mass model in the time domain. Thus, if \( * \) denotes convolution, then

\[
p_{1}(t) = z_{in}(t) * u_{s}(t)
\]

and similarly for \( p_{L} \) and \( p_{N} \).

By utilizing the relationship between input impedance and input reflectance Schumacher [12] has shown that an alternative form for \( p_{1} \) is

\[
p_{1}(t) = Z_{0} u_{s}(t) + r_{m}(t) * [p_{1}(t) + Z_{0} u_{s}(t)],
\]

where \( Z_{0} \) is the characteristic impedance of the first section of the vocal tract model and \( r_{m}(t) \) is the inverse Fourier transform of the input reflectance of the tract, \( R_{in} \). This recursive representation is better (especially for computing the pressure build-up during oral closure) because \( r_{m} \) has a much shorter effective duration than does \( z_{in} \).

3.2.2 Aspiration and Friction. For aspiration we follow the suggestions of Flanagan [Ref. 13, p. 251]. The squared Reynolds number is calculated corresponding to the glottal flow \( u_{s} \). Then if \( Re^{2} \) is greater than a critical value, \( Re^{2}_{\text{crit}} \), a uniform random noise given by

\[
p_{ng} = \varepsilon_{ng} \text{random}(Re^{2} - Re^{2}_{\text{crit}}),
\]

is substituted for \( p_{ng} \) in Eq. (1). Here \( \varepsilon_{ng} \) is an empirically determined gain (about \( 2 \times 10^{-6} \)), \text{random} \ is a random number uniformly distributed between -0.5 and 0.5, and \( Re^{2}_{\text{crit}} \) is empirically found to be about 2700. For friction generated in the vocal tract we follow Flanagan's suggestions [14] but with several modifications. First, we introduce friction at the narrowest constriction only rather than at every point of the tract. Second, we use a volume velocity noise source rather than a series pressure source. This is because the pressure source cannot be placed at the constriction (because the internal impedance of the source becomes too high). Instead, the source has to be placed downstream of the constriction at different distances for different stops and fricatives. Shadle [15] has shown that for a volume velocity source the placement is not so critical, and we position it one section downstream of the constriction. The strength of the volume velocity noise source is computed as \( u_{s} = p_{d} / R_{s} \) where \( R_{s} \) is the constriction resistance and \( p_{d} \) is the noise pressure source controlled by the smoothed (low-passed to about 2000 Hz) flow through the constriction. The strength of \( p_{d} \) is computed by a formula similar to Eq. (8), except that \( Re^{2}_{\text{crit}} \) is 202 and the gain .002.

4. SUMMARY

In this paper we have outlined a description of an articulatory speech synthesizer in which the properties of the oral and nasal tracts are computed in the frequency domain, converted to the time domain by inverse Fourier transformation, and interfaced with a time-domain nonlinear model of the vocal cords. The synthesizer allows for nasalization, frication, and aspiration. With all these features included the program runs at about 5 times real time on a Cray-I computer, when the sampling rate is 20 kHz.

REFERENCES

A METHOD FOR CONNECTING SPEECH SYNTHESIS UNITS USING SPECTRAL TRANSITION MATRICES

T. Saito, Y. Matsuda and T. Kaneko

Science Institute, IBM Japan, Ltd.
5-19, Sanban-cho, Chiyoda-ku, Tokyo 102, Japan

INTRODUCTION

In speech synthesis by rule, it is important to smoothly connect synthesis units (e.g. phonemes, syllables etc.) in order to obtain high quality synthesized speech. Most works reported previously employ a simple interpolation of the spectral parameters such as linear interpolation for two adjacent synthesis units. However, unit to unit transition is so complex that substantially more information is needed for smoother connection.

In this paper, we propose a new method for connecting speech synthesis units. The information of the inter-unit transition is defined as a spectral transition matrix and used for connecting units. In case of Japanese CV-syllables, we need about 100 matrices. Such spectral transition matrices are obtained from actually observed data. Experimental results showed that this method gives such smoother transition closer to continuous speech.

SYNTHESIS UNIT

SPECTRAL PARAMETER

As a spectral parameter of synthesis unit, LSP (Line Spectrum Pair) is used for this study. LSP parameter is one of the mathematically equivalent representations of LPC spectral information, and it is well-known to have lots of properties superior to other parameters ([1]) such as: (1) filter stability; (2) quantization; (3) parameter interpolation etc. (3) is especially important to speech synthesis by rule from the point of unit connection.

CV-Unit

Japanese speech can be expressed by a combination of only about 100 CV-syllables (Consonant and Vowel pair) and the number is relatively small compared with other languages. Besides, Japanese speech is said to have a "syllable-isochronous rhythm." It means that, in a series of syllables, each syllable's duration does not change drastically. Through using CV-syllables as synthesis units in Japanese synthesis, we have the following advantages: (1) The number of synthesis units is small; (2) The prosody control rule could be simplified. From these points of view, we use CV-syllables as synthesis units.

A CV-unit is segmented as shown in Fig.1. The segmented unit is connected with a preceding vowel and evaluated by listening test. The beginning point of T1 is determined based on listening test. The end point of T2 is chosen such that the vowel stationary part is about 100 ms long.

Spectral Transition Matrix for VC Transition

When connecting two adjacent CV-units, it is critical how well the transition of connecting part can be generated. In previously reported works ([2,3,4], simple interpolation models (such as two or three point linear interpolation) have been used to the connection. However, such connections are not smooth enough and the synthesized speech lacks naturalness.

In order to improve the connection, it is important to observe the actual transition of syllable to syllable in continuous speech and to obtain the information through this observation. In this paper, we introduce -- SPECTRAL TRANSITION MATRIX for VC transition -- to improve the connection of CV-units. The matrix is extracted from the actual transition of vowel to consonant, and it keeps only the differential values of spectral parameters during the period (T3,T4) (see, Fig.2). Transition from vowel to the following consonant begins at the point T3. T4 is usually selected around the consonant onset point depending on each consonant feature.

A spectral transition matrix \( t(i,j) \) is defined using LSP parameters of \( w(i,j) \) as follows.

\[
 t(i,j) = w(i,j) - w(i-1,j) \\
 (i = 1, \ldots, N, \ j = 1, \ldots, p)
\]

where \( N \) is the number of frames for the transition period (T3,T4), and \( p \) is the analysis order of LSP parameters.

CONNECTING OPERATION

At first, the duration of each syllable is determined from the prosody control information which is obtained through text analysis. Then, connecting operation for adjacent CV-units is carried out by bridging CV-units using a spectral transition matrix. Smoothing technique is applied at the two connection boundaries. The duration of each syllable is controlled by lengthening or cutting the stationary part of vowel in CV-unit.

CV-Unit Bridging

At the first step of connecting operation, two adjacent CV-units are bridged by inserting another unit generated from a matrix for the VC transition. The matrix can not be inserted as it is because the actual parameter values of unit boundaries are different in each case. Therefore, it is necessary to generate another unit that has the same boundaries as CV-units on both sides. Two bridging functions \( W(j) \) and \( W*(j) \) are defined as follows.

\[
 W(j) - W*(j) = \frac{w(j,j) - w(0,j)}{w(N,j) - w(0,j)}
\]

where \( j = 1, \ldots, N \).
\[ W2(i,j) = \left(1 - \frac{1}{N} \right) f(i,j) + \frac{1}{N} g(i,j) \]

where

\[ f(i,j) = \sum_{k=1}^{p} x(k,j) \]

\[ g(i,j) = W2(i,j) = \sum_{k=1}^{p} x(k,j) \]

Wv(j) : LSP parameters at the last frame of the preceding CV-unit

Wc(j) : LSP parameters at the first frame of the following CV-unit

Function (B1) has the possibility to emphasize the spectral transition too large for drastic transition. On the contrary, function (B2) does not have this property. Therefore, function (B2) is considered to be more suitable for CV-unit bridging. Fig. 3 shows CV-unit bridging by function (B2).

Boundary Smoothing

As the spectral transition of the vowel part in a CV-unit is almost flat, the generated transition may have a sharp angle at the connection boundary if only the bridging function (B2) is carried out. To avoid this phenomena, a simple smoothing technique is applied for the boundaries. This smoothing is carried out by using the average value of two adjacent parameters at the boundaries.

EXPERIMENT

Experiments were carried out for a Japanese phrase of /watashiwa/ ("I" in English) + a particle (wa). The speech was sampled at 10 kHz and LSP parameters (analysis order p = 10) were analyzed at every 10 ms with 30 ms Hamming window.

The spectral parameters of the phrase were obtained by connecting 4 CV-units of /wa/, /ta/, /sh/ and /wa/. The connection is performed using the proposed method and the conventional linear interpolation. For the proposed method, the needed spectral transition matrices for the connections of /a/ -> /t/ , /a/ -> /s/ , and /i/ -> /a/ were obtained from VCV utterances of /ata/ , /as/ and /iwa/, respectively.

Fig. 4, 5 and 6 show the transitions of LSP parameters for the phrase. Fig. 4 and 5 show the case of linear interpolation and proposed connecting method, respectively, for the case of natural utterance for the phrase. So, we find that the proposed connecting method gives much smoother transition closer to continuous speech than the conventional linear interpolation. In the transition of /j/ to /wa/, distinctive formant transitions are observed, and they are critical factors in the sense of perception. The linear interpolation method cannot represent the transitions. Using the proposed method, it is possible to represent a rather complex

cated transition such as /l/ to /w/.

CONCLUSIONS

We proposed a new method for connecting speech synthesis units. This connecting method is composed of CV-unit bridging technique and boundary smoothing technique. The information for the actual inter-unit transition is stored as a spectral transition matrix and used for the connecting operation. Experimental results showed that the connecting method gives much smoother transition closer to continuous speech than the conventional linear interpolation method.

REFERENCES


OUTILS DE LA THEORIE DE L'INFORMATION

Soit y, un symbole et X, son contexte, c'est à dire une chaîne de 1 ou 2 symboles le précédent. Nous appelons Hxy l'entropie conditionnelle de y connaissant x (quantité d'information des séquences de 2 ou 3 symboles) et HXy le l'entropie conditionnelle de y connaissant x (quantité d'information apportée par y quand on connaît son contexte x). Les manuels de théorie de l'information contiennent des formules pour calculer les grandeurs à partir de fréquences des suites de 1, 2 et 3 entités. Elles sont reliées par la relation

Hxy = Hx + Hyx

La redondance au niveau d'un symbole y est

R = 1 - (Hyx / HO),

où HO est le nombre de bits qu'il faudrait pour coder les symboles s'ils étaient équiprobables.

OPPOSITIONS DE GRANDES CLASSES PHONETIQUES

Pour évaluer l'importance respective des oppositions de grandes classes phonétiques, notre approche consiste à neutraliser une opposition donnée, et à évaluer quantitativement l'impact de cette neutralisation sur certaines grandeurs caractéristiques. Nous avons choisi de procéder ainsi d'une part au niveau d'un dictionnaire de cohortes, et d'autre part au niveau des entropies sur les suites de classes.

Toute hiérarchie entre les oppositions de classes doit être comparée à celle qui est introduite "naturellement" par les fréquences d'occurrences (classement en fonction de la probabilité conjointe du couple (Pipj)) ou du taux de couverture du texte (Pipj) si Pi et Pj sont les fréquences des classes i et j). Dans le cas présent, cette classification place en tête des couples (V, autre), en raison de la fréquence de la classe V, puis ensuite les couples de classes consonantiques fréquentes, et enfin de classes consonantiques rares.

Dictionnaire phonétique et cohortes

Nous avons obtenu, à partir de notre corpus, un dictionnaire de mots phonétiques, qui compte 7221 entrées, avec leurs nombres d'occurrences (102317 occurrences de mots dans le corpus). Ce dictionnaire est publié (/6/) et commenté (/7/) par ailleurs.

Nous utilisons maintenant, comme indicateur, le nombre de phonétes formées de chaque classe par la classe phonétique à laquelle il appartient. On obtient un dictionnaire de "cohortes" pour lequel le nombre de total de cohortes, le nombre de celles ne comportant qu'un seul mot ("unicines") et la taille moyenne (au niveau dictionnaire et au niveau corpus) et la taille maximale des cohortes constituent autant d'indicateurs de l'efficacité d'un système de classes pour l'accès au lexique (/8/). Voici ces chiffres pour les systèmes à 8, 6 et 4 classes définis ci-dessus.

<table>
<thead>
<tr>
<th></th>
<th>original 8 cl.</th>
<th>6 cl.</th>
<th>4 cl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>cohortes</td>
<td>7221</td>
<td>3946</td>
<td>2463</td>
</tr>
<tr>
<td># occurrences</td>
<td>12.8</td>
<td>25.9</td>
<td>29.6</td>
</tr>
<tr>
<td>entropie</td>
<td>6.9</td>
<td>7.1</td>
<td>6.7</td>
</tr>
<tr>
<td># uniques</td>
<td>7221</td>
<td>2856</td>
<td>2463</td>
</tr>
<tr>
<td>% diction.</td>
<td>100%</td>
<td>39.5%</td>
<td>34.1%</td>
</tr>
<tr>
<td>% corpus</td>
<td>100%</td>
<td>8.3%</td>
<td>7.5%</td>
</tr>
<tr>
<td>taille max</td>
<td>1</td>
<td>30</td>
<td>64</td>
</tr>
<tr>
<td>taille moy(dico)</td>
<td>1.0</td>
<td>2.1</td>
<td>2.5</td>
</tr>
<tr>
<td>taille moy(corp)</td>
<td>1.27</td>
<td>19.9</td>
<td>24.2</td>
</tr>
</tbody>
</table>

Si l'on confond maintenant deux classes, l'importance des variations au niveau de ces cohortes fournit un critère pour évaluer l'importance de leur opposition. C'est ce que nous avons fait, pour le système à 8 classes.
Pour toutes les grandeurs envisagées, les résultats sont concordants: les oppositions les plus efficaces (celles dont la suppression dégrade le plus l'accès au lexique) sont celles entre classes consonantiques fréquentes: (L, O-), (F-, O-), (O-, O-), (N, O-) etc., dans cet ordre. Les oppositions (voyelle, classe consonantique) se situent en bas de la liste (la première, (V, L1) étant vers le milieu). Ce fait traduit bien entendu l'importance de la structure dominante CVVCV... dans le message.

Variations d'entropie

Une autre approche consiste à calculer l'entropie conditionnelle 

\[ \text{Biyx} \] au niveau classe phonétique, et la variation de cette grandeur lorsqu'on neutralise l'opposition de deux classes. Les plus fortes diminutions de cette quantité d'information sont enregistrées pour les oppositions de classes consonantiques fréquentes: (LI, O-), (LI, F-), ... Au contraire, pour les oppositions (voyelle, consonne), l'entropie conditionnelle augmente: il est plus difficile de prédire le symbole suivant si la structure CV est partiellement détruite.

Note: nous avons également appliqué cette méthode de traitement au niveau des oppositions de phones. On observe les mêmes tendances, les oppositions (consonne, consonne) et accessoirement (voyelle, voyelle) ne diffèrent nettement de celles du type (consonne, voyelle); mais les variations d'entropie observées sont faibles, et on ne peut être sur qu'elles soient significatives.

TRAITS DISTINCTIFS

On peut classer les traits distinctifs en fonction de leur entropie, c'est-à-dire du degré d'indépendance des contextes de ce trait. Voici une idée a priori de leur efficacité potentielle. Voici ce tableau (pourcentages + et - , et entropie):

<table>
<thead>
<tr>
<th>Trait</th>
<th>Entropie</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF</td>
<td>57.3</td>
</tr>
<tr>
<td>INT</td>
<td>38.0</td>
</tr>
<tr>
<td>AIG</td>
<td>40.6</td>
</tr>
<tr>
<td>ROC</td>
<td>77.5</td>
</tr>
<tr>
<td>BEM</td>
<td>23.6</td>
</tr>
<tr>
<td>VOT</td>
<td>87.5</td>
</tr>
<tr>
<td>EXT</td>
<td>7.4</td>
</tr>
<tr>
<td>TRA</td>
<td>6.3</td>
</tr>
<tr>
<td>NAS</td>
<td>13.5</td>
</tr>
</tbody>
</table>

Mais on tient mieux compte de la structure du langage en procédant de la façon suivante: si l'on supprime un trait (c'est-à-dire si l'on neutralise son opposition +/-), on crée des confusions de phones, qui à leur tour engendrent des cohortes de mots dans le dictionnaire phonétique. Les mesures définies précédemment sur ce dictionnaire de cohortes permettent d'apprécier l'importance du trait ainsi étudié. Voici le tableau correspondant, ordonné d'après la taille moyenne de cohorte dans le corpus:

<table>
<thead>
<tr>
<th>Trait</th>
<th>Nombre de cohortes</th>
<th>Taille moyenne de cohorte</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIF</td>
<td>10</td>
<td>6293</td>
</tr>
<tr>
<td>AIG</td>
<td>11</td>
<td>6734</td>
</tr>
<tr>
<td>ROC</td>
<td>7086</td>
<td>668.8</td>
</tr>
<tr>
<td>BEM</td>
<td>7100</td>
<td>6777</td>
</tr>
<tr>
<td>VOT</td>
<td>7057</td>
<td>6777</td>
</tr>
<tr>
<td>EXT</td>
<td>7025</td>
<td>6777</td>
</tr>
<tr>
<td>TRA</td>
<td>7119</td>
<td>6777</td>
</tr>
<tr>
<td>NAS</td>
<td>7121</td>
<td>6777</td>
</tr>
</tbody>
</table>

Cette classification met en valeur l'importance des traits dont la suppression crée des confusions voyelle - voyelle ou consonne - consonne (EXT, BEM, AIG, NAS, VOT), et d'inverses celles des traits qui diffèrent des paires voyelle - voyelle (VOC, TRA, INT).

CONCLUSION

Il apparaît que les oppositions les plus porteurs d'informations, donc celles que des systèmes de décodage acoustique - phonétique devraient tenir à prendre en compte les premières sont les suivantes:

- pour les classes phonétiques, les oppositions de classes consonantiques fréquentes
- pour les traits distinctifs, ceux qui, à la fois pertinents pour les consonnes et les voyelles, permettent des distinctions entre voyelles, et entre consonnes (Diffus, Aigu, Bémoléen).

Ces résultats sont bien entendu en lien avec la forte structure CVCV... observée dans la langue parlée.

BIBLIOGRAPHIE

/5/ P. FONSALE: Connected word recognition system using speaker-independent phonetic features. ICASSP 85, pp 312-315
/7/ J.P. TUBACH, L.J. BOE: Quantitative knowledge on word structure, from a phonetic corpus, with application to large vocabularies recognition systems. ICASSP 82 pp 546-549.
POURQUOI UNE BASE DE DONNEES DES SONS DE PAROLE ?

Les recherches en traitement automatique de la parole présentent aujourd'hui deux problèmes d'activité principaux : l'évaluation des algorithmes de reconnaissance, et l'étude de nouveaux modèles en reconnaissance et en synthèse de la parole.

Aujourd'hui, s'il existe des bases de données lexicales en documentation automatique, en traduction assistée par ordinateur et en reconnaissance de la parole pour la plupart des langues, les bases de données des sons ne font que l'appréhension que depuis quelques années seulement à cause de leur complexité de mise en œuvre : choix des corpus et des locuteurs, volume de stockage important et modalités d'échange et de stockage. Ces bases de données constituent les premiers éléments de ce que l'on commence à appeler les "industries de la langue".

Depuis 1983, 15 équipes regroupées au sein du GRECO (Groupe de Recherche Coordonnée) CNRS "Communication Parlée" mettent en place une telle base pour la langue française. Le volume et la diversité des données manipulées (signal acoustique, étiquettes et résultats d'analyse), a rendu indispensable l'utilisation d'un système informatisé de gestion et d'accès.

I. CHOIX DES CORPS ET DES LOCUTEURS

Ne pouvant couvrir l'étendue des besoins de tous les laboratoires représentés, nous avons dû nous limiter, dans une première étape, aux deux sous-ensembles suivants :

- (A) le corpus "Évaluation", enregistré par 32 locuteurs. Il est composé de 3 parties :
  - le corpus AE : 5 phrases titulées "la bise et le soleil" et 54 groupes CVCC comportant les trois voyelles /a/, /i/, /u/ dans les environnements consonantiques extrêmes /p/, /s/, /h/ en phrase porteuse.
  - Chiffres et Nombres : 400 chiffres isolés, 200 suites de 3 chiffres, 100 suites de 2 chiffres et 100 suites de 5 chiffres "connectés", ainsi que 100 nombres de 0 à 99 et 100 nombres de téléphone.
  - Lettres et Noms : 432 lettres de l'alphabet isolées, 52 noms épelés "en isolé" et 50 noms épelés "en continu".

- (B) le corpus "Acoustic", enregistré par 12 locuteurs. Il est composé de 2 parties :
  - de Mots : 600 groupements CVVC comportant les 17 consonnes du français et les 3 voyelles cardinales /a/, /i/, /u/, 200 groupes consonantiques, et les Tests de Rimes (1) pour les consonnes et les voyelles par paires et par triplés.
  - de Phrases : 50 phrases phonétiquement équilibrées (2), 44 phrases pour l'étude des nasales et 192 phrases contenant des mots réels avec toutes les consonnes et toutes les voyelles du français.

Le choix des locuteurs, en dehors de la norme de prononciation, a été le suivant : 12 locuteurs de base ont enregistré les corpus (A) et (B). Ils ont été sélectionnés par un ensemble de 6 phonéticiens comme étant représentatifs de la prononciation "standard" du français. Les 20 locuteurs complémentaires, qui ont enregistré le corpus (A), sont répartis ainsi : 10 représentent divers accents régionaux et 10 ont été choisis sur la base de leur élocution particulière qui les rend difficiles à reconnaître par les systèmes automatiques. Il y a autant de locuteurs féminins que de masculins.

II. ENREGISTREMENTS, FICHIERS, SUPPORTS

Les enregistrements sont directement numérisés sur 16 bits à une fréquence d'échantillonnage de
16 kHz (basse passante : 7,5 kHz) au CNET de Lannion.
Le fichier complété a un rapport signal/bruit de 85 dB. Les éléments à prononcer sont présentés sur une console de visualisation et synchronisés avec l'acquisition de façon identique pour tous les locuteurs.

Le logiciel d'enregistrement et d'archivage des bandes fonctionne sur PDP11 sous RT11. Trois studios répondant aux mêmes normes, seront installés à partir de l'année 86 dans d'autres laboratoires du GRECO.
Les fichiers ont une structure qui constitue une norme nationale. Ils sont composés d'un ensemble de blocs contenant toutes les informations nécessaires à l'accès aux différentes parties de l'enregistrement, du signal numérique lui-même, et d'une fin de fichier image de listing détaillé accompagnant celui-ci. En outre, ceux-ci sont compatibles avec ILS.

Depuis le mois de juillet 1985, l'ensemble des enregistrements est stocké sur 320 bandes magnétiques numériques.
- Le corpus (A) occupe un volume de 125 mégaoctets par locuteur, c'est à dire un total de 2,3 Giga-octets (soit environ 20 heures d'enregistrement pour les 32 locuteurs).
- Le corpus (B) occupe un volume de 125 mégaoctets par locuteur, c'est à dire un total de 2,3 Giga-octets (soit environ 12 heures d'enregistrement pour les 12 locuteurs).

Pour diffuser cette base, un nouveau support d'archivage a été choisi : l'enregistrement numérique sur cassettes vidéo au standard bétabox.
Le GRECO a fait mettre en place, dans 12 laboratoires français, un équipement constitué d'un convertisseur Sony PVM (ou T21), associé à un enregistreur vidéo Bétabox et connecté sur plusieurs ordinateurs - hôtes possibles grâce à l'équipement OROS - AI (société ORS de Grenoble). Celui-ci permet l'échange de données numériques entre calculateur et cassette. De plus, ce système fait, en temps réel, le changement de fréquence de la valeur standard 44,1 kHz (pour l'Europe, PAL ou SECAM) des enregistrements contenus dans la cassette, à une valeur quelconque comprise entre 5 et 44 kHz (éventuellement en stéréophonie) tout en conservant la qualité originale. Enfin, ce dispositif possède sur la deuxième piste un "marquage" d'informations en ASCII pour les entêtes et définissant le contenu de la bande (directory) ainsi que des informations temporaires. L'information OROS - AI est actuellement disponible sur PDP et LSI 11 sous RT11, sur IBM PC sous MS-DOS (via une carte spécialisée contenus dans le PC) et sur VAX sous VMS.
Ce nouveau type de support est extrêmement économique : une cassette de 2 heures peut contenir environ 230 Mocents, soit l'équivalent de 18 bandes magnétiques 1600 bpi (2400 pieds). Une fois dupliqué, le corpus (A), sera contenu dans 10 cassettes, et le corpus (B), dans 8 cassettes seulement. Les transferts se feront à l'aide d'un automate de cette année, et toutes les équipes pourront accéder à la base avant le milieu de l'année 86. Ainsi, nous disposons...
d'un moyen d'"archivage multi - machines standardisés" pour tous les laboratoires permettant un échange simple et économique d'enregistrements effectués à n'importe quelle fréquence.

III. SEGMENTATION ET ETIQUETAGE

La segmentation et l'étiquetage constituent une étape très importante dans l'élaboration de la base de données des sons du français (BDSONS) et de qualité des opérations menées à ce stade, dépendra la richesse de la consultation ultérieure. Le GRECO a choisi de disposer à la fois d'un étiquetage dit "large", utile pour l'interrogation de la base et le repérage des réalisations en contexte phonémique large, et d'un étiquetage dit "fin" destiné plus particulièrement à l'étude fondamentale des sons du français.

Un groupe de travail a défini les principes de la segmentation et de l'étiquetage (3) qui seront appliqués. L'étiquetage "large" consiste à repérer et à étiqueter les centres des réalisations de segments de taille phonémique alors que l'étiquetage "fin" correspond à un repérage d'événements qui traduisent des discontinuités majeures (une dizaine environ) apparaissant sur le signal de parole et/ou sur son spectre. Pour l'étiquetage "fin" deux méthodes complémentaires sont mises en oeuvre :

- la première se situe dans le domaine temporel et opère directement sur l'onde acoustique; c'est donc une méthode entièrement manuelle.
- la seconde qui est une méthode semi-automatique, procède à partir d'une représentation fréquentielle.

Ce qui est réalisé d'espérer pouvoir monter l'étiquetage "large" entièrement automatique, mais l'évaluation des diverses méthodes possibles et le choix de l'un d'elles reste à faire.

Une structure à plusieurs niveaux (étiquetage hiérarchisé) sera superposée à cet ensemble pour attribuer chaque son aux deux discontinuités successives, à une de macro-classes parmi la dizaine actuellement définies.

Une évaluation approximative du temps nécessaire à l'étiquetage "fin" des groupes CVCV du corpus du type Annostique est de l'ordre de 1 homme-année. Cette tâche sera entreprise en 66.

IV. STRUCTURE ET ACCÈS À LA BDSONS

L'objectif principal du Système de Gestion de la Base de Données (SBDB) est de permettre l'accès aux informations enregistrées en facilitant leur recherche selon certains critères formulés à la consultation. Un SIGB relationnel muni d'un langage d'interrogation interactif et d'un interface programmable facilitant son intégration à divers programmes d'application a été utilisé.

L'importance du volume de données à gérer (de l'ordre de 4 Mega-octets) nous a conduit à procéder en deux étapes :

1. Gestion des descripteurs des fichiers-sons.
   Dans un premier temps, nous gérions les descripteurs des fichiers-sons (stockés hors de la base sur bandeaux ou cassette magnétiques). Descripteurs contiennent des références aux supports qui permettront l'accès postérieur au corpus souhaité.
   Chaque enregistrement peut être recherché à travers des requêtes sélectives qui portent sur les caractéristiques des locuteurs, des conditions d'enregistrement, des corpus prononcés et sur la description de leur contenu.

2. Gestion des sons proprement dits.
   Dans un deuxième temps, nous envisageons la gestion détaillée du contenu des fichiers-sons en exploitant les étiquetages "large" et "fin" tels qu'ils ont été définis au paragraphe III.

Pour cette étape, il sera nécessaire d'accéder aléatoirement aux fichiers-sons en les stockant sur un support de type "Gigadisc". Les descripteurs et les résultats de l'étiquetage seront conservés sur disque magnétique.

A ce niveau, il sera possible d'obtenir des copies de certains corpus de la BDSONS sur des cassettes avec leurs fichiers d'étiquetage associés, enregistrées. Pour la première étape, la structure fondamentale de la BDSONS s'appuie sur 4 entités de base qui correspondent à des relations du schéma conceptuel : LOCATEUR, CORPUS, REALISATION et ELEMENT.

LOCATEUR qui précise son nom, prénom, sexe, âge, carnet-linguistique, adresse, tel
CORPUS (code-corpus, orga-corpus, nb-éléms/corpus)
REALISATION (code-corpus, code-locuteur, support, no-unité, début-réalisation, durée, date, lieu, heure)
ELEMENT (code-corpus, code-locuteur, type-éléms, nb-éléms-real, début-éléms, durée)

Pour permettre la consultation par description des éléments, on ajoute la relation :
CONTEXTE-CORPUS (code-corpus, no-éléms, description-phonétique)

Ce schéma conceptuel permet d'évaluer le volume de données à gérer à environ 3 Mega-octets.

La consultation de la BDSONS s'effectue à travers des requêtes portant sur celles relations. Par exemple :
"Donner les descripteurs des réalisations du premier corpus de syllabes indiquant pour chacune ; support, no-unité, début-réalisation, durée et les locuteurs qui ont prononcé ces syllabes, qui exploitent les résultats des sons pour accéder aux réalisations existent sur les cassettes PUM et pour les copier sur un autre support en vue du traitement.

Une version 0 de la première étape de la BDSONS est opérationnelle depuis janvier 86.

CONCLUSIONS

Un travail fondamental sur la Langue Française est amorcé. La BDSONS est la base d'un consensus des ingénieurs et des phonéticiens français. De nombreuses réunions de concertation ont abouti à un accord sur la méthodologie adoptée pour définir les caractéristiques des enregistrements, le choix des locuteurs et des corpus, l'étiquetage et la segmentation. Des programmes normalisés d'analyse ont aussi été retenus. La deuxième phase du développement de la BDSONS commence : l'étiquetage "large" et "fin", distribution à la demande de corpus particuliers, intégration des résultats d'analyse ont aussi été retenus. La deuxième phase du développement de la BDSONS commence : l'étiquetage "large" et "fin", distribution à la demande de corpus particuliers, intégration des résultats d'analyse ont aussi été retenus.

Par ailleurs, il a été proposé d'interconnecter BDSONS et BDLEX (également soutenu par le GRECO Parole). Ainsi sera ouvert un champ important : "du son jusqu'aux mots".

Références bibliographiques :

AUTOMATIC LABELING SYSTEM FOR LARGE SPEECH DATABASE

Shozo Makino and Hisashi Wakita

Speech Technology Laboratory
3888 State Street,
Santa Barbara, CA, 93105, USA

INTRODUCTION

A large amount of segmented and labeled speech data with corresponding phonetic transcriptions are essential for developing a speech recognition system based on phonetic units, as well as for developing a text-to-speech synthesis system. Traditional manual labeling is extremely time-consuming and subject to lack of consistency and reproducibility of the results.

Over the past few years, several automatic time alignment procedures have been proposed in the literature. Most of these approaches attempt to align the input utterance with a manually labeled reference utterance using dynamic time-warping[1,2]. A second approach, which also uses dynamic programming, is to segment and label the utterance into broad phonetic classes independent of speakers prior to time alignment[3,4]. Although broad phonetic classes are robust, much better labeling results are expected if a system is speaker-dependent and the reference patterns for phonetic units can be made from input speech.

This paper describes an approach to such a labeling system. The labeling system is composed of two parts. The first part makes reference patterns for 73 phonetic units from 70 CVC words. The second part segments arbitrary word utterances uttered by the same speaker into phonetic units. Using the training reference patterns and voiced-consonant information, the input speech is first aligned with the transcription using dynamic programming with duration constraints for each phonetic unit, and then more accurate phonetic boundaries are determined using new reference patterns derived from the input speech.

EXTRACTION OF REFERENCE PATTERNS

73 phonetic units derived from 50 phonemes are used for the labeling. 48 of the 50 phonemes are included in ARPA-BET. Two other phonemes are unvoiced vowels. Each voiceless stop and voiced stop in the 50 phonemes has three phonetic units, corresponding to buzz portion, burst portion and segmental portion. Each diphthong has two phonetic units, corresponding to the first half portion and the remaining portion. 70 CVC words contain most word-initial and word-final consonants and stressed vowels. The remaining phonetic units which cannot be extracted from the 70 CVC words are substituted by similar phonetic units. An end point is detected using the logarithmic power and the zero crossing rate. Vocalic-consonant decision is made using the parameters with Laplacian filter computed for each frame. Going back from the frame with maximum power to the beginning frame of the speech, the first consonant frame is regarded as the final frame of the word-initial consonant segment. Going forward from the frame with maximum power to the end frame of the speech, the first consonant frame is regarded as the beginning frame of the word-final consonant segment. If the vowel is a single vowel, a typical frame is extracted from the vowel segment using the frame with maximum power. If the vowel is a diphthong, the frame with the nearest distance from a similar single vowel is regarded as the typical frame. Accordingly the two typical frames are extracted from the vowel segment for a diphthong. A typical frame for the word-initial consonant is extracted from the segment between the beginning frame of the speech and the typical frame of the vowel segment using the parameter shown in the table. A typical frame for the word-final consonant is extracted in a similar fashion. The labeling of the 70 CVC words begins with the words which can be easily segmented into necessary phonetic units. In the process of labeling, the reference patterns of phonetic units are temporarily generated and used for obtaining typical frames. Figure 1 shows an example of the labeling for /m ah ch/. 'B' and 'E' indicate the beginning and end frames of the speech. The frame with maximum power is the 32-nd frame. The typical frame for /m/ is defined as the 5-th frame after the beginning frame of the speech, that is, the 19-th frame. The frame(28-th) with equal distances from both typical frames is regarded as the boundary. The frame(53-th) with maximum derivative of the logarithmic power(\(C_0\)) is regarded as the burst frame of /eh/. The typical frame for the buzz portion of /eh/ is defined as the 3-rd frame before the burst frame. The frame(43-rd) with equal distances from the typical frames of the vowel and the buzz portion is regarded as the boundary.

After the labeling of the 70 CVC words, reference patterns for each unit are computed by means of vector quantization clustering[6] from the speech samples which are gathered up the 5 frames around the typical frame of the corresponding phonetic unit. At most 8 reference patterns per unit are extracted, where a reference pattern is represented by 10 cepstral coefficients.

AUTOMATIC LABELING SYSTEM

The end-point detection algorithm generates at most two candidates for the beginning frame and for the end frame of the speech. The beginning frame is determined by the maximum power of the logarithmic power. The end frame is chosen from the frames which are the nearest distance from both typical frames of the vowel and the frame of the consonant.

Table 1 Parameters for extraction of typical frames and boundaries in consonants

<table>
<thead>
<tr>
<th>Phoneme</th>
<th>Typical frame</th>
<th>Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>group</td>
<td>Initial</td>
<td>Final</td>
</tr>
<tr>
<td>y, w, r, l</td>
<td>/i/</td>
<td>/a/</td>
</tr>
<tr>
<td>m, n, m</td>
<td>/a/</td>
<td>/a/</td>
</tr>
<tr>
<td>p, k, ch, jh</td>
<td>max</td>
<td>max (dC_0)</td>
</tr>
<tr>
<td>b, d, g, dx</td>
<td>max (dC_0)</td>
<td>min (dC_0)</td>
</tr>
<tr>
<td>h, l, s, ah</td>
<td>min (C_1)</td>
<td>min (C_1)</td>
</tr>
<tr>
<td>v, y, sh, s, sh</td>
<td>/a/</td>
<td>/a/</td>
</tr>
</tbody>
</table>

Fig. 1 An example of labeling for /m ah ch/
frame of the speech. The end-point is finally determined using continuous DP[6].

The input speech whose end-points have been detected is then time-aligned with its phonetic description using dynamic programming with duration constraints[7]. A reference template is made by concatenation of the phonetic units with its maximum duration. The distance between the reference frame and the input frame is defined as the weighted distance which is computed by the distance with weighting coefficient. The distance is defined as the minimum distance in the M distances to the input frame computed by M-reference pattern of phonetic unit corresponding to the reference frame. The weighting coefficient is dependent on vocalic-consonant information of the input speech frame and the phonetic unit of the reference frame. The beginning and the final frames of each segment are discarded by going back from the reference frames to the optimal path.

Using the phonetic segment information obtained from the duration-constraints DP, the reference pattern of the phonetic unit is renewed. The three-frame averaged distance is the same as the one-frame averaged distance computed for each frame in the segment. The three frame averaged cepstral coefficients around the minimum distance are regarded as the new reference pattern of the phonetic unit. The distances from the new reference pattern are recomputed from the beginning frame of the preceding segment to the final frame of the following segment. Based on the recomputed distances, the duration-constraints DP is applied again. The beginning and the final frames of each segment are also defined by going back from the end frame along the optimal path.

LABELING EXPERIMENTS

Five methods of automatic labeling were evaluated on 15 repetitions of 104 words uttered by two males and one female.

In method 1, input speech is time-aligned with the corresponding word reference template in the first repetition uttered by the same speaker using a dynamic time warping method[8]. In method 2, the 73 phonetic units and duration constraints are used for the labeling without vocalic-consonant information. In methods 3 through 5, vocalic-consonant information is used for the automatic labeling. In addition, methods 4 and 5 employ adaptation to the input speech, that is, new reference patterns for the phonetic units are derived from the input speech using the results obtained by method 3. Methods 1 through 4 use the manually extracted end-points, whereas method 5 uses the automatically extracted end-points.

Four types of reference patterns were also tested. The first type (Ref 1) is a single pattern per phonetic unit, manually extracted from the 104 word vocabulary. The other types are multiple patterns per phonetic unit and are computed by means of vector quantization clustering. The second type (Ref 2) is computed from speech samples extracted from manually labeled typical frames in the 104 key words. The third type (Ref 3) is computed from speech samples extracted from manually labeled typical frames in the 70 CVC words. The fourth (Ref 4) is computed from speech samples which are derived from automatically extracted typical frames in the 70 CVC words.

Table 2 shows the comparison between the labeling methods. The values in the table represents the averaged difference between the method's standard deviations. The standard deviations of the distributions of differences between manually and automatically labeled boundaries for 104 words were averaged over the 15 repetitions and the difference taken. The comparison between methods 2 and 3 shows that the vocalic-consonant information remarkably improves the accuracy. The comparison between methods 3 and 4 shows that the adaptation to the input speech also improves the accuracy. Method 4 which gave the best results is almost equal to method 1 based on word reference patterns. However, method 5 indicated rather poor results due to insufficient end-points detection.

Table 3 shows the comparison between reference patterns. Use of multiple reference patterns per phonetic unit is very effective. Ref 3 gave nearly the same results as those obtained by Ref 2 except the female voice. This indicates that the 70 CVC words contain sufficient information for making multiple reference patterns per unit. The large difference for the female is due to the fact that the 104 key-board vocabulary was recorded 3 years earlier than the 70 CVC words, and, in addition, the subject had a slight cold when the 70 CVC words were recorded. Automatic extraction of typical frame from the 70 CVC words works well as indicated by the fact that the results obtained by Ref 4 were nearly the same as those obtained by Ref 3.

Table 2 Comparison between labeling methods(unit: ms)

<table>
<thead>
<tr>
<th>Method</th>
<th>Type of reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref 1</td>
<td>Ref 2</td>
</tr>
<tr>
<td>Method 2 - Method 3</td>
<td>2.6</td>
</tr>
<tr>
<td>Method 3 - Method 4</td>
<td>4.8</td>
</tr>
<tr>
<td>Method 4 - Method 5</td>
<td>-3.5</td>
</tr>
<tr>
<td>Method 1 - Method 4</td>
<td>-2.0</td>
</tr>
<tr>
<td>Method 1 - Method 5</td>
<td>-2.7</td>
</tr>
</tbody>
</table>

Table 3 Comparison between reference patterns(unit: ms)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Sub.</th>
<th>Type of method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref 1 - Ref 2</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>Ref 2 - Ref 3</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>Ref 2 - Ref 3</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>Ref 3 - Ref 4</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>Ref 3 - Ref 4</td>
<td>F</td>
<td>1</td>
</tr>
</tbody>
</table>

REFERENCE

POUR DECRIVER EFFICACEMENT LES PHONETISMES.

R. Thomas, Peymans, 63170 Brignoles, France.

QUESTIONS PRÉLIMINAIRES AUX ACOUTISTIENS.

A l'inverse de la plupart des acousticiens qui Mighty un petit nombre, ou un seul des paramètres des sons, les enseignants de langues étrangères doivent décrire ces sons assez complètement pour permettre de les interpréter et de les reproduire. C'est pourquoi il nous faut vous demander, à vous les spécialistes, si nous notons sur la parole et notre façon de les présenter concordent avec les connaissances actuelles.

Nous réitérons donc d'abord notre question fondamentale, restée sans réponse bien qu'elle importe autant à vos recherches qu'à l'étude des idiomes étrangers. Peut-on reconnaître une position neutre de l'appareil vocal pour la parole ? Autrement dit, par exemple: Dans la parole, lorsqu'il n'y a pas d'exigence particulière telle que la production des sons voisés, les cordes vocales sont-elles plus tendues, ou plus fermées, que normalement dans la respiration?

Pardonnez l'imprécision de cette question correspondant à des déclarations opposées sur ce sujet par M. P. Lieberman(1) et R. Husson(2) ; il paraît aujourd'hui que cette opposition vient de ce que la réponse semble varier selon les langues et qu'aux ne précisent pas qu'ils ne parlent que d'une seule et limitant les formulations à la langue propre à chacun d'eux, Lieberman et Husson semblent avoir tous deux raison, le premier soutenant alors que le larynx est généralement ouvert dans la parole anglaise comme dans la respiration, et le second trouvant qu'à l'inverse du comportement habituel, le larynx est fermé le plus souvent possible en prononciation française.

Il nous faut aussi rappeler que P. Delattre(3) a donné cette double réponse que nous adoptons aussi parce qu'il semble y trouver une clé au problème séculaire de millions d'étudiants et professeurs d'anglais et de français langue étrangère. Car, depuis 29 ans, sans que nous soyons convaincus qu'au contraire des laboratoires rhénans l'ont aussi vérifié, nous avons constaté qu'en changeant, par des moyens pratiques, cette attitude fondamentale dans la parole, les étudiants anglaisphones et francophones peuvent parler l'autre langue sans accent particulier(4).

QUESTION PRINCIPALE: COMMENT DECRIVER UN PHONETISME?

Mais, venant maintenant au cœur de notre propos, en tenant compte de cette opposition dans un des principaux mécanismes de la parole, on arrive à surmonter un obstacle infranchissable autrement, doit-on pour cela dire qu'on trouve dans ce comportement du larynx la différence fondamentale entre les prononciations anglaise et française ? Non, puisque l'efficacité pédagogique n'est qu'un indice, si frappant qu'il soit. Tout ce qu'on peut affirmer, c'est qu'il est possible, mais non certain, que ce soit là la cause principale des différences entre ces deux phonétismes. Même du simple point de vue du contraste entre deux langues, il n'est pas aussi facile de résoudre le problème d'une description efficace des prononciations. C'est cependant possible, et comme on le verra, une telle possibilité est loin d'être négligeable.

Notons en effet d'abord son importance dans l'enseignement des langues étrangères : on doit enseigner une à une, et trop souvent sans grande efficacité, des différences qui, avec une bonne description des phonétismes, seraient bien appréciées d'un seul coup.

Ensuite, cette carence semble encore plus gênante pour la recherche en acoustique de la parole. En effet, cela revient à dire qu'il existe des différences entre les sons des diverses langues, mais souvent nous ne savons que faire à leur sujet. Comme l'écrit J. Leaer, entre autres: "les racines intellectuelles de la phonétique moderne se nourrissaient d'une culture anthropographique d'une influence alphabétique extrêmement grande.

Nous savons que l'alphabet et ce qu'on peut dire de ses utilisations ne constituent pas une description suffisante de nos phonétismes, mais comment compléter celle-ci ?

PREMIÈRES DEMARCHES POUR UNE DESCRIPTION NOUVELLE.

Il faut d'abord reconnaître ouvertement que continuer à partir des bases est de plus en plus inacceptable, à mesure que va diminuer le volume de travail accumulé. Nous devons oser dire ensuite qu'il nous est de moins en moins possible d'avoir une vue d'ensemble cohérente de phonème pourtant très flexibles du langage, dans les prononciations évidentes et de manière commune, nous ne pouvons faire un autre choix. C'est l'ouverture de notre science présente, qui ne peut donner cette description générale, ni ne la pourra si nous n'organisons pas nos recherches autrement que nous ne le faisons.

Soulier l'insuffisance de nos études présentes.

Pour être sûr de ne pas omettre d'éléments importants, revoions d'abord ce que nous savons de plus général sur cette activité, c'est-à-dire la philosophie et la psychologie de la parole. Les dernières autorités universitaires actuelles, mentionnées par exemple K. Kant, Russsou et Homsky, reconnaissent qu'on ne peut définir le langage, notre objet d'étude, que par la prononciation. Leur objection nous a conduit dans cette indication négative. Certes, en 1969, M. Polanyi(5) a proposé une délimitation que l'on ne refuse pas et qui paraît très pertinente : mais il en découle des développements, ils s'intercalent dans notre phrase, pour utiliser l'expression de l'auteur du signe.

On ne saurait plus attendre une telle description des études plus propres et linguistiques, si précieux que puissent être leurs fruits par ailleurs. La psychologie a fait des remarques et, encore après P. Delattre, nous reconnaissions l'utilité pédagogique de certaines d'entre elles : mais on ne saurait trouver dans de tels études un moyen de décrire avec une prédiction succédant l'ensemble des prononciations. Les mécanismes neurophysiologiques sont souvent mentionnés avec précision et cependant les problèmes en sujets restent nombreux, même dans des domaines aussi élémentaires que fondamentaux que celui concerné par notre question préliminaire. Surtout, on ne connaît que de façon bien imparfaite plusieurs des hiérarchies et dépendances qui lient ces mécanismes entre eux. Plus gênant encore, on ne sait pas toujours quels effets les mouvements qu'ils causent produisent sur les sons du langage. Si ce sont les idées qui se suivent sans mécanismes de mémoire, on ne peut donc pas attendre de ce côté non plus une prédiction définie de la parole ou des prononciations.

C'est dans l'étude acoustique de la chaine parlée que les triomphes semblent les plus nombreux. Les analyse et statistiques d'une profondeur et d'une précision imparable, obtenues permettent maintenant la synthèse mécanique de la parole et promettent une amélioration remarquable des méthodes auditives(7). Mais c'est aussi ici qu'on sait le mieux qu'il est
vain de compter trouver une réponse proche à nos doutes généraux. C'est encore en ce domaine qu'on note le plus clairement la différence qu'il y a entre les traitements mécaniques de la parole, traitements dont nous pourrons de mieux en mieux contrôler les étapes reconnus, et ceux qui font moins d'effets. Il nous faut donc s'entendre sur certaines parties, l'indécente diversité. Et qui pourrait publier que ce sont ces traitements faits "spontanément" par l'homme lui-même, tout en ils sont pratiquement inconnus, sont plus efficaces?

INFORMATIONS COMPLÉMENTAIRES NON ENCORE UTILISÉES

Mais, ce survol de nos connaissances et recherches présentes, si négatif qu'il soit directement pour notre propre d'une meilleure description de nos phonétiques, est par là même d'une grande utilité: il montre combien des points de vue nouveaux peuvent être souhaitables pour mieux organiser des recherches et que les données existantes sont de plus en plus nombreuses. On a ébréché des ensembles spéciaux de l'étude, chez des malades par exemple, de langages déficients dans leur production ou leur réception. Pourquoi laisser inexplorée l'énorme domaine que constituent les expériences de millions de malades et d'étudiants de langues étrangères?

Depuis des siècles, on a eu recours à toutes sortes de procédés et de règles pour aider à l'adaptation de phonétiques étrangères, tout en conservant ceux de nos langues qui se sont écartées de la diversité des langues est telle qu'on n'a pas encore classé les différences reconnues entre les prononciations, ni à plus forte raison, les moyens employés pour surmonter ces écarts.

Ce savoir est tout à fait distinct de celui que la recherche phonétique utilise et un peut entendre deux choses de son examen par les acoustiques de la parole. D'une part le silence de ces derniers doit permettre d'améliorer beaucoup cet acquis traditionnel, au moins en éliminant les contradictions qui ont émis des procédés. Vous m'avez dit autre part, notre question préalable suffirait au besoin à prouver que l'examen approfondi de ces méthodes, dont pour certaines la valeur est indiscutable, pourraient apporter des points de vue nouveaux sur les questions les plus importantes. L'essentiel naturellement global dans beaucoup de cas serait contre, au moins, un examen partiellement, à la dispersion pratiquement imposée par les spécialisations présentes.

CE QUI EST IMMÉDIATEMENT SOUHAITABLE.

Notre propos ici ne peut être de commencer vraiment cette étude, mais d'attirer notre attention sur cette source si abondante en informations d'ensemble sur le fonctionnement de la parole, faire un inventaire et un classement aussi utile que possible de tous les procédés en usage pour passer d'un phonétique à l'autre, cherchant en même temps à éviter d'oublier chacun d'eux, tel semblerait devoir être le premier goal d'une grande association qu'il nous faudrait créer, peut-être sous l'égide d'un groupe existant déjà.

Il ne serait certainement utile à l'étude des langues étrangères, mais pourrait être aussi précieux pour une description prochaine beaucoup plus exacte des divers phonétiques avec l'éclairement ainsi l'un l'autre. Peut-être cela conduirait-il même ensuite à des définitions plus générales encore.

NOTES

3/ Delattre Pratique de phonétique française, Middlebury College, Middlebury Vermont, 2e éd., 1951. Sans être explicites, ses réponses y sont seulement implications. Il est notable que nos deux chercheurs ne semblent pas avoir eu une connaissance des faits sur lesquels Delattre s'appuie, faits d'intérêt surtout pédagogiques.
5/ J. Lawer The phonetic description of voice quality Cambridge UP 1990 p.7 "The intellectual roots of modern phonetics: are embedded in an orthographic culture whose alphabetic influence is very pervasive" (our translation).
6/ Dans "Sense-Giving and Sense-Reading" Philosophy Macmillian Oct. 1987. Polanyi y a publié une leçonnaire dont il savait qu'elle allait à l'encontre des idées qu'il avait soutenues; il n'est abattu d'en parler ensuite. Voir notre thèse et cet article parut par après.
7/Tous ces faits auxquels nous ne pouvons faire mention sont bien connus. Pour le dernier cas, celui des études pour les promesses auditives, on peut voir l'article de J. Gémin dans "Les Journées d'Études sur la Parole" CNET Paris 1965 pp. 10, 13, 79...

BIBLIOGRAPHIE

Les méthodes de correction phonétique ne sont pourtant nullement décrites dans des publications. Quelques ouvrages généraux, nous indiquent quelques ouvrages permettant une connaissance un peu plus précise du sujet.

W. Brooks Language Learning Harcourt 1964
Chukski N. Language and Mind NY Harcourt 1969.
Delattre P. "De la Hiérarchie des indices acoustiques pour la perception de la parole" Proc. 8th IC Phoc. Sciences 1962 244-253.
Gilet A. Gymnastique phonétique franco-anglaise Paris Didier 1990.
Green H. Elements of English Phonology Paris PUF 1962
"Teaching Foreign Pronunciations" Abstract in JASA Suppl.1 Vol 76 Fall 1974 Minneapolis.
EFFICIENT TRAINING OF NON-NATIVE SPEECH DISCRIMINATIONS WITH PERCEPTUAL FADING.

David E. Morosan and Donald G. Jamieson

Speech and Audition Laboratory, Department of Psychology, University of Calgary, 2500 University Dr. N.W., Calgary, Alberta, T2N 1N4, CANADA

Linguistic experience produces major changes in the perception of speech sounds. For example, English speakers place the phonetic boundary separating /b/ and /p/ at approximately 25 ms voice-onset time (VOT; cf., Lisker and Abramson, 1970; Williams, 1977), while Spanish speakers place this boundary at approximately 0 ms VOT (cf., Williams, 1977). Similarly, the Canadian French voiced/voiceless categorical boundary occurs at a shorter VOT value than does the unilingual English boundary (cf., Carmazza et al. 1973).

While distinctions appropriate to a language group are easily learned in childhood, some non-native speech contrasts appear to be remarkably difficult for adults to learn (e.g., Werker and Teras, 1984). It seems likely that at least some failures to train non-native contrasts reflect deficiencies in the specific training method. We believe that training: 1. should ensure that the relevant speech cues are presented in an acoustic context which is appropriate for normal speech, rather than in isolation; 2. should involve classification, if the desired outcome is to improve speech sound identification; 3. should first use extreme contrasts to focus attention on critical cues, and then progress to approximate the natural language environment. We applied these principles to train a contrast which appears remarkably difficult for many Francophones to acquire -- the English /β/ vs /s/ contrast.

Method

We used a variation of the perceptual fading technique (Terrace, 1963), with synthetic prototypes as exemplars. During training, we systematically increased the amount of intra-phonemic variation by adding new stimuli, one at a time. These new sounds were progressively less salient exemplars of the voiced and voiceless tokens. Stimuli were synthesized in cascade at a 10 kHz sample rate, using Klatt's (1980) cascade/parallel

four sounds in each stimulus set (i.e., voiced or voiceless) differed in the duration of frication, decreasing from 140 ms to 35 ms of frication in 35 ms steps, from the most extreme to the medial tokens of the continuum. Sixteen natural speech tokens were selected from productions by a single speaker, on the basis of their variability in frication duration. All stimuli were presented binaurally to subjects at level of 70 dB over AKG 240 headphones.

Twenty bilingual Canadian Francophones were grouped into ten matched pairs of subjects, on the basis of their error rates in identifying the natural speech tokens. All subjects continued to participate in their English immersion course throughout the experiment. Control subjects received no experimental training. Subjects in the training group received identification training consisting of a series of trials in which one synthetic token was presented at a time, for the subject to identify by pressing one of two labelled microswitches. Twelve training tapes were used, beginning with the easy task of identifying the most extreme stimuli of the continuum (i.e., 1 and 8). For subsequent tapes, the task became more difficult as more medial stimuli from the continuum were introduced into the set of stimuli to be identified. Thus, the number of stimuli presented in each tape increased from the two most extreme stimuli (1 and 8) to the entire set of eight sounds over the first 10 tapes.

Results and Discussion

To allow changes in sensitivity to be measured independently of such biases, identification responses were converted to A' scores (McNicol, 1972) using each subject's hit rate with a given stimulus, in combination with that subject's overall error rate on all stimuli of the opposite type (e.g., "voiced" responses with voiceless stimuli) as the False Alarm rate. The figure shows that while training improved listeners' identification of both voiced and voiceless synthetic tokens (F(1, 9) = 32.00, p < .01), the control group did not improve from pretest to posttest (F(1, 9) = 1.47, p > .05). For the natural tokens, A' also increased after training (F(1, 9) = 8.75, p < .05) but not for control subjects (F(1, 9) = .35, p > .05). Thus, with both natural and synthetic stimuli, training improved identifications of voiced and voiceless fricatives.
Comparison of pre-test and post-test identification performance for voiced and voiceless synthesized sounds for listeners in the control and training groups. Each point represents the mean of 40 A' scores, collapsed across the four tokens within a stimulus set and the ten listeners within each group.

We tested the hypothesis that training would increase linguistically-relevant discrimination by comparing the pre- and post-test A' values for each stimulus pair. A one-tailed dependent t-test showed that the A' score increased after training (t(9) = 2.43, p < .02 for the pair 4-5). No other discrimination score, for either group, showed a significant change in A' scores over time. Thus, training was effective in improving performance on the most difficult discrimination between voiced and voiceless fricatives, but no change occurred for the control group. Importantly, training did not change within-category discrimination. Since accurate speech perception requires subjects to make accurate discriminations between phonemes on the basis of minimal cues, while failing to discriminate between allophones on the basis of other (even large) non-phonemic acoustic differences, this result demonstrates that the training task was successful in improving discrimination in a linguistically meaningful fashion.

**General Discussion**

Our results are encouraging for attempts to train adult listeners to identify and discriminate non-native speech sounds. Just 90 minutes of practice in identifying synthetic speech tokens from a structured continuum improves the identification of both synthetic and natural speech sounds, and inter-category discrimination, while leaving intra-category discrimination unchanged. Performance improved for both voiced and voiceless stimulus types and for both synthetic and natural speech. Finally, since the control group showed remarkable consistency from pretest to posttest, the performance improvements shown by the trained subjects cannot be attributed to factors such as improved performance under the experimental conditions.

*This work was supported by grants from the AHFMR, HWC, and NSERC to D.G. and by an AHFMR Scholarship to D.M. We are grateful to Ms. E. Rogers for providing testing facilities in Kingston and for helping to obtain subjects, and to Ms. C. McDermid for her assistance in providing subjects at Calgary.*

**References**


EFFECT OF AUDITORY DELAY ON AUDIO-VISUAL SPEECH PERCEPTION

P.C. Pandey, S.M. Abel, and H. Kunov

Institute of Biomedical Engineering, University of Toronto, Toronto, Ont., Canada M5S-1A6.

INTRODUCTION

In recent years, there has been a considerable interest in developing cochlear prostheses, for partially restoring the auditory function by direct electrical stimulation of the auditory nerve [1-5]. As only limited information can be transmitted through these devices, it is imperative that the important features of speech be extracted and properly encoded on the stimulating electrodes. Digital signal processing methods can be employed for the extraction of these features. Knowing the delay parameters, one can get a reliable estimate of feature parameters in a particular pitch-period, the values in the next two or more periods have to be observed [4]. A certain amount of additional delay will be involved in the processing. It is desirable that the auditory information should complement the visual information available through lip-reading, used by many deaf individuals [5], and the auditory delay introduced in the feature extraction/encoding should not disrupt lip-reading.

Several studies on the effects of auditory delays have been reported. Maximum disruptive effects of delayed auditory feedback on normal hearing adults' speech occur for delays of 200 ms [6]. Dixon and Spitz noted formal lip-reading experience, auditory delay detectable during audio-visual presentation of connected speech (broadband, unprocessed) as 257 ms [7]. However, it is possible that even if the delay is not detectable, the desynchronization between auditory and visual signals may cause phonemic confusions [8]. McGrath and Summerfield studied the effect of auditory delay with the voice fundamental frequency (from a laryngograph) used as the auditory signal in conjunction with lip-reading by normal hearing adults [9]. They report that for a delay of 160 ms, the performance goes down to that of lip-reading, and that the actual amount of delay is unimportant so long as it is less than 80 ms.

In the experiment reported here, we were interested in the effect of auditory delays when the auditory signal carries more information than just the fundamental pitch, but is not sufficient by itself for unaided speech perception. It can be hypothesized that individuals with a background in lip-reading make more efficient use of visual information and therefore may be more susceptible to auditory delays as compared to naive subjects. To investigate this, we carried out two experiments. Subjects in Experiment I did not have formal lip-reading experience, while subjects in Experiment II had some earlier exposure to it, either as part of their professional training or through association with the deaf.

EXPERIMENT I

Method

Video-taped lists of sentences from the lip-reading test of the Iowa Cochlear Implant Battery [10] were used as the test material. There were two lists, each comprising 30 sentences. In one list, each sentence is preceded by a picture of an object, which would be in the sentence. This list was designated "with context" (C). The other list did not have context information and was designated "without context" (NC). The speaker was a male, speaking un-accented American English. Twelve subjects (six male and six female, 19-30 years age) with normal hearing (speech reception threshold in 0-5 dB HL range) and normal (or corrected) vision participated in the study. They had no previous training in lip-reading or any exposure to the test material. The test was carried out in an acoustically isolated room. The acoustic signal was presented monaurally to the subject's right ear (to the left ear in the case of a left-handed subject) through a pair of Telephonics TDH-49P headphones. The video signal was presented on a 20-in color monitor. The distance between the subject and screen was 1.3 m.

The audio track signal from the recording was passed through a digital delay line (implemented on a microcomputer, IBM-PC with TEGAM Labmaster card), with 10 kHz sampling rate and 12-bit amplitude resolution. The delayed auditory signal was mixed with a 60 dB multi-talker babble mask noise. There were two S/N ratios (0 and -10 dB) and six delays (0, 60, 120, 180, 240, and 300 ms).

Each sentence list was subdivided into 6 sub-lists (each with 5 sentences, containing 23 to 39 words). Subjects participated in three 30 minute experimental sessions. During each session, all the twelve (six C and six NC) sublists were presented. Of the 18 sublist presentations for each kind of list, the first three were used for lip-reading practice and the others for the 15 test conditions. A randomization scheme was used to counterbalance the presentation order of context, delay and S/N ratio combinations across subjects.

The subjects were instructed to repeat the sentences (or those words they had perceived), paying attention to the tense and plurality markers. Two minutes of score was recorded in each of the correct words (23-29 per sublist) and total number of correct key/context words (13-16 per sublist).

Results

There was virtually no difference in the results of the two scoring methods. The zero delay audio-visual scores for all subjects were better than their respective lip-reading or hearing scores. The 300 ms delay, no subject's scores were better than those with normal hearing adults [9]. They report that for a delay of 160 ms, the performance goes down to that of lip-reading, and that the actual amount of delay is unimportant so long as it is less than 80 ms.

In the study reported here, we were interested in the effect of auditory delays when the auditory signal carries more information than just the fundamental pitch, but is not sufficient by itself for unaided speech perception. It can be hypothesized that individuals with a background in lip-reading make more efficient use of visual information and therefore may be more susceptible to auditory delays as compared to naive subjects. To investigate this, we carried out two experiments. Subjects in Experiment I did not have formal lip-reading experience, while subjects in Experiment II had some earlier exposure to it, either as part of their professional training or through association with the deaf.

EXPERIMENT II

Method

Six subjects (two male and four female, 26-47 years age), with normal hearing (speech reception threshold in 0-10 dB HL range) and normal (or corrected) vision, participated. All subjects had some familiarity with lip-reading either as part of professional training or through working/association with the deaf. No attempt other than the experiment itself was employed to quantify their lip-reading ability. The experimental procedure and the instructions given to the subjects were the same as
Experiment I. Only four delay conditions (0, 80, 160, and 240 ms) were used. To further shorten the time required, only one S/N condition, -5 dB, was used. Each sublist (5 sentences) was presented under one condition. The order of context and delay conditions were randomized across the subjects.

Results
The two scoring methods gave similar results in this experiment. Repeated measures analysis of variance on audio-visual scores showed significant effects of main factors (context) and (S/N) but no significant effect of 2-way interaction. Fig. 2 shows the pooled (C and NC) group mean total word scores. The general nature of the disruptive effect of the delay is the same as in experiment A. There is no significant difference between lip-reading and 240 ms delay audio-visual scores.

A comparison of Figs.1 and 2 indicates that subjects with lip-reading experience may be more susceptible to the disruptive effects of auditory delay. However, the evidence is not clear cut.

CONCLUSION
The first experiment employed normal hearing subjects without any formal experience in lip-reading. The auditory information was degraded by a masking babble noise. The disruptive effect of the delay turns out to be a function of both the context and the S/N ratio, becoming more dominant when S/N is low. The effect of delay was more clearly visible in the second experiment, for subjects with a lip-reading background. The findings that delays up to 80 ms do not affect the scores are in agreement with the theory that sensitivity to audio-visual desynchrony is not significant for phonemic identification in connected speech, and that the effect of the delay is important at a syllabic level [9].

In terms of implications of these results for speech processing for cochlear implants, we conclude that if the information provided is high, moderate delays of up to 80 ms (which is more than 4 pitch periods) introduced in processing and encoding will not outweigh the benefits of auditory signal. Of course, these findings are based on experiments with normal hearing subjects only. It should be mentioned that the delay will also affect the auditory feedback to the patient of his own voice. Again, if the related studies on normals [6] are taken as a guide, delays of two to four pitch periods should not have a disruptive effect.

ACKNOWLEDGEMENTS
The work reported here was in part supported by MRC program grant # PG-11 (IRK), NSERC grant # A4478 (SMA), and DND/DCEM grant # 9711-1-4-897 (SMA). We thank Ms Grace Chua for her assistance, the Canadian Hearing Society for the help extended by its staff, and the individuals who participated as subjects.

REFERENCES
DISPLA YING SPEECH FOR DEEP DEAFS

G. Vilac la ra

Ecole Polytechnique Fédérale de Lausanne, 16 ch. de Bellerive, 1007 Lausanne, Switzerland.

INTRODUCTION

This paper presents a novel form of display, which reproduces the speech signal in a coordinate system, where the main axes carry information relevant to the learning of speech by the deep deaf children. Presently the absence of a simple visual representation, which would allow the deaf and dumb child to control his voice, is the main gap of his apprenticeship. Nowadays the sole instruments used by speech therapists are pitch-meters, sound-level-meters, spectro-meters and even oscilloscopes.

When observing a steady voiced sound on the oscilloscope, its periodicity is interpreted as the repetition of a pattern characteristic of the pronounced sound. The signal can be segmented in perfectly superposable elements called pitch periods. By modifying the melodic component, a deformation by anamorphosis with no great alteration of the pattern is observed. During the transition between two distinct, voiced sounds the former pattern changes its shape progressively to become the new one. On the other hand unvoiced sounds have to be considered in a completely different way. They must be taken as entities, whether they are occlusives, noises or even silences. Any segmentation performed over such signals is arbitrary.

THE MODEL USED BY THE ALGORITHM

The model, on which the algorithm is based, is defined by the following hypothesis:
- The speech signal is constituted by a succession of alternatively voiced and unvoiced segments.
- A segment is voiced if it can be subdivided into look-alike subsegments called pitch periods.
- Two successive pitch periods are linked through a relationship of similarity.
- The time ratio between two pitch periods (similarity ratio) must not exceed 30%.
- This similarity is not perfect. A maximal amplitude variation of 30% is accepted between two successive pitch periods brought by similarity to the same duration (this relation is called an anamorphosis).

DESCRIPTION OF THE ALGORITHM

The algorithm operates directly on the sampled speech signal p(k). The relative amplitude of the fundamental harmonic is increased using a nonlinear operator \( p(k) = p(k) \) for \( p > 0 \) and \( p(k) = p(k)/2 \) for \( p > 0 \) [1]. The signal is then band-pass filtered. The upper cutoff frequency corresponds to the maximum frequency be detected and the lower one allows to remove the DC component introduced by the nonlinear operator.

Meanwhile each filtered sample is compared to the preceding one in order to generate a binary signal that localizes all the zero crossings having a positive slope ZCPSs. These particular points subdivide the speech signal into time sections. Segments are then constituted by placing side by side successive time sections. We define an hypothetical pair of pitch periods, by gathering two successive segments, which duration difference does not exceed 30% (the duration is given by the number of samples embodied). Each pair contains a possible pitch period followed by a candidate to comparison.

After having defined all the pairs, beginning at the starting point (ZCPSs(0)), each one is verified by comparing the global shape of the two possible periods composing it. For this purpose it is necessary to reduce both of them to a normalised duration in order to allow a direct comparison by superposition. The normalizing process is carried out by subdividing each possible period into 16 subsegments evenly distributed. As long as the number of samples, composing each period, is a multiple of sixteen (e.g. \( N = 32 \)) each subsegment can be defined as \( "X" \) successive samples. "X" being the number of samples divided by sixteen \( (X = N/16 \leq 2) \). Otherwise (e.g. \( N = 31 \)) the subsegments chosen by this method will only represent a truncated part of the possible period \( (X = 31/16 = 1) \).

A more suitable procedure is to subdivide each possible period into 16 subsegments, which length can be single or double. If the number of samples is \( N = 16X^2 \) (e.g. \( N = 16 \times 15 \times 15 = 31 \)), the segmentation process generates "U" subsegments including "X" samples and "V" subsegments including 2"X" samples according to the relation \( N = (U+2V)X^2 \) (31 = (1+2+15)1).

The latter method is equivalent to a nonlinear subsampling, which rate is adapted to the length of the analyzed possible period. Its global shape is obtained in a normalized form, by sharing uniformly and in the right proportions the large subsegments among the little ones. The flowchart, on figure 1, shows the segmentation process for any period regardless of its length.

---

**figure 1:** Flowchart of the segmentation process

Sixteen elements, given by the average value of the samples included in each subsegment represent a possible period. The pair that scores the shorter average distance between the corresponding elements is considered as a pair of pitch periods as long as the relative distance does not exceed 30% (hypothesis assumed by the model). Although two pitch periods are identified at a time, the process is carried on at the beginning of the second period. The detection continues using the preceding period of comparison as the only possible period. The latter is kept, even if this detection fails. In this case the research process is reset. If there is a gap, before the next period is identified, a choice is presented: either the gap length is comparable to the neighbour periods length and it is considered as a period or the gap is held as an unvoiced segment. This means that the detection of one pitch period out of three is enough to segment the whole signal.
DEFINITION OF THE DETECTION LIMITS

First of all we have to keep in mind that successive subsampling processes affect the spectrum, drawing up the spectral images to the base band. Aliasing is avoided by limiting the detection to a frequency interval of two octaves and choosing an appropriate filter. Considering the greatest pitch period, that can be detected by the algorithm, we define the length of the analysis window, in such a way that it encloses two successive pitch periods of this particular type. If the shortest period is 16 samples in length, the longest one will embody 64 samples, which means a window's length of 128 samples.

If the detection range has to be larger, the research process must take place separately in two different intervals. Supposing a sample frequency of 15 kHz and a low-pass, cutoff frequency of 937.5 Hz, the resulting signal allows a pitch detection within the interval [234.4, 937.5 Hz]. Subsampling this signal by a factor of 4 (375 Hz) and submitting it to a second identical filter (the cutoff frequency is shifted down to 234.4 Hz), the final signal allows a second detection within the interval [58.9, 234.4 Hz].

If the length of the possible period is in the threshold zone between the two intervals ([234.4 Hz]), an incommene may occur: the pair to be analyzed might be too long and exceed the upper window ([234.4, 937.5 Hz]) and one of its possible periods too short to be represented by 16 samples. In this case the detection has to be made in the upper interval. This means that the second possible period (the candidate to comparison) will be truncated. Therefore the first one must be truncated in the same proportions in order to be compared to the second one.

REDUCING INFORMATION FOR REAL TIME PROCESSORS

The farther the end of a period is, from the starting point (i.e. the longer the period is), the more samples will be included in a subsegment. Imposing this number of samples as being a power of two, it is indeed possible to define a code as a succession of predefined subsegments, which length increases with the distance from the origin.

Therefore with a code composed of 46 subsegments including one sample, 23 including 2 samples and 9 including 4 samples, we reduce, from 128 to 78, the number of samples used by the process.

Using such a code, the amount of data transmitted to the processing unit (NEC µPD-7720) is minimized. This is important because the internal data RAM of this processor, specialized for signal processing in real time, is limited to 128 words [2].

THE DISPLAY OF SEGMENTED SPEECH

Indeed the segmented signal calls for a three-dimensional representation. Pitch periods are disposed one behind the other in a chronological order, whereas the unvoiced sounds are arbitrarily segmented (cf figure 3).

figure 3: Representation of the french word "exercise"

CONCLUSION

By means of the algorithm proposed in this paper it becomes possible to segment automatically the speech signal. The display shows clearly the evolution of the pitch. It helps to distinguish most consonant labial doubles [3]. This technique is quite promising for the learning of speech by the deep deaf children.

ACKNOWLEDGEMENTS

It is to be reminded, that the idea of such a display was first proposed by Henry Eden [4], who uses an accelerometer disposed on the throat in order to detect the glottal pulses. These are used to trigger the segmentation. By means of the algorithm proposed in this paper it becomes possible to operate without a separate trigger signal.

The author thanks the "Fonds National de la Recherche Scientifique Suisse", for its contribution to this research.

REFERENCES


SPEECH TRAINING ON A PERSONAL COMPUTER

H. Crépy, F. Destombes, M. El Béze, G. Rouquié

IBM France Scientific Center, 36 avenue Raymond Poincaré, 75116 - Paris (FRANCE) - (1) 45 05 14 00

Giving the deaf child a visual feedback to help him control his vocal productions is quite an old idea, as soon as 1946, Bell Labs introduced the "Visible Speech" which displayed spectrograms onto a spinning CRT [1]. This was a fully analog device.

Later on, many other electronic devices and some computers were used to the same purpose [2,3,4,5,6,7]. Moreover, some systems can be found for sale which display pitch or intensity curves on a TV screen or an oscilloscope.

Technically speaking, most of these systems use a combination of two technologies:

- speech signal acoustic analysis performed by analog electronic circuitry,
- display and control using a mini- or micro-computer.

In 1978, the IBM France Scientific Center started a research project on this subject. A first prototype (with 8086 microprocessors) was installed at Institut National des Jeunes Sourd (Paris) at the end of 1979. The only analog part of this system dealt with speech signal input. All functions were implemented in a microcomputer with computation, as well as display on a black & white TV screen. This technical decision for a fully programmed system already allowed for developing voice-controlled games.

However, this system, although programmed, was not really "programmable" in the assembly-language sense. Programs were stored in ROM, and the user could not program new applications. When the IBM Personal Computer (PC) was introduced in 1981, it was decided to use it as the central component of a new speech training system, fully programmable.

IBM PROGRAMMABLE SYSTEM: HARDWARE

The system uses two main parts: an IBM-PC and a prototype speech input and processing board connected to a microphone. It is thus built from standard hardware, except for the prototype speech board. Besides, the PC may also be used for other purposes aside from speech training sessions.

IBM Personal Computer

The minimum configuration includes:

- 1 IBM-PC with 256 Kbytes RAM
- 1 diskette drive (double sided, double density)
- 1 color/graphics adapter and display
- 1 keyboard

Other IBM Personal Computer models may be used (PC/XT, PC/AT...). While only one diskette drive is required, a second one (or a hard disk) will be found handy for file editing, program development, etc. Also, a graphics printer will be needed should one want to keep a hard copy, e.g. of a pitch curve produced by a child.

Prototype speech processing board

This board was designed at the IBM France Paris Scientific Center, and a small series (100 replicas) was manufactured at the IBM plant of Bordeaux. It is important to note that this board is a prototype and, therefore, 100 boards are available. The schools for the deaf participating in the experiments have received it either as a loan or as a gift. Nearly the whole batch of 100 boards is now allocated, and there is no plans for any more production.

From a technical point of view, this board consists of the following elements:

- 80186 general purpose microprocessor (8MHz)
- 128 Kbytes RAM
- 32 Kbytes ROM
- microphone input circuitry: filter, sampler (up to 15kHz), 12 bits analog/digital converter. It fits as any other expansion board into one of the option slots of the PC.

IBM PROGRAMMABLE SYSTEM: SOFTWARE

The speech board has its own microprocessor, thus two programs are running simultaneously in the system:

- The speech board is in charge of the voice input sampling and the repetitive power-consuming acoustic processing. These programs are stored in the on-board ROM, but other programs, or modified versions can also be loaded and run from the on-board RAM. The object-code compatibility with the PC (80186/8086) makes it quite easy to develop those programs.
- The PC assumes all the control and display functions, plus more processing if required. The PC programs are generally written in BASIC, and make extensive use of the graphic capacities of the PC.

A speech therapist with sufficient knowledge of BASIC may therefore design new programs to visualize speech features. This is the key interest of this programmable system.

Speech board software

Without going into too much details, let us simply note that the programs in the speech board ROM allow it to process in real-time one (and only one) of the following parameter sets computations:

1. Pitch (fundamental frequency of the vocal chords), intensity, amplitude, zero-crossings (rate of changes in the signal sign).
2. Intensity, zero-crossings, differentiation, Hamming window, multiplication, 11 normalized autocorrelation coefficients.
3. Pitch, intensity, amplitude, zero-crossings, differentiation, multiplication, 5 normalized autocorrelation coefficients.

The chosen parameter set is computed every 128-sample frame (i.e. every 12.5ms, or 80 times per second). These parameters are commonly used in speech processing applications.

When dealing with deaf children speech training, they can be used as follows:

- Parameters of set (1) are relevant to supra-segmental or prosodic features (pitch, intensity, rhythm, rhyme), and can therefore be used in early training programs.
- Set (2) allows for a more sophisticated analysis (e.g. vowel recognition), and can thus help addressing problems later raised during articulation training.

Basic PC application package

This is the set of PC speech training programs which is provided as a basis for experiments. It is expected to grow through user programming (see next section).

- Program selection menu

Allows an end-user with limited computer familiarity to run the programs quite easily.

- Amplitude adjustment

Displays two bar-graphs (intensity in decibels, amplitude in number of significant bits), along with instructions on how to adjust the pre-amplifier to input a correct level signal (and avoid saturation).

- Pitch and/or intensity graphs

Basis of the first stages of training, shows graphs in real-time onto two half-screens (one for the teacher's pattern, and one for the pupil's tries). Real-time visual feedback of a simple sentence, the length of a vocal production, the voiced/unvoiced discrimination between some phonemes (e.g. /f/ vs. /v/, /s/ vs. /z/), or /p/ vs. /b/m/.

- Pitch animated games

The goal here is to try to lead a mobile towards targets, while avoiding obstacles (e.g. a camel towards pond, a sheep towards grass, a fox doing the running for cheese, but escaping from cats etc. The vertical position of the mobile is proportional to
the pitch within an user-selected range. Various preset targets and obstacles layouts and an interactive screen layout definition mode are available for varying difficulty.

- **Voicing controlled game**
  A target the teacher has designed a mountain/valley pattern, the child tries to drive a balloon across the screen, flying over the mountains by voicing, while allowed to get his breath back when crossing a valley.

- **Vowel articulation game**
  A mobile is led through a maze by pronouncing four vowels, one for each of the four moving directions. Many options allow for varying the difficulty level of the game (maze complexity, speed...), and choosing the four vowels group among those that the program can effectively recognize.

- **Voicing and amplitude graph**
  Two graphs (teacher/pupil) can be displayed in real-time. Their thickness shows the signal amplitude, while their color (red or green) indicates the voiced/unvoiced distinction.

- **Face**
  A different representation of the same speech features is a semi-animated face whose mouth size is proportional to amplitude, while the eyes color change with voicing.

- **Sentence parameters display**
  This program displays parameters of the (1) and (2) steps for a 3 seconds sentence. It can also show the sampled signal waveform at a given instant in that sentence. This display is too complex to be used by children, but may be of interest for deaf adults or teenagers, as well as for speech therapists. It is also a useful research tool for preliminary studies before a new program design.

  All the above programs take their messages from external files. Translation to new fel of languages is therefore very easy. Currently available are French, English, German, Spanish, Italian, Dutch, Hebrew and Arabic versions.

**User programs development toolkit**

The required software consists of:

- **DOS** (Disk Operating System) version 2.0 up,
- **BASIC** compiler (version 1.0) or
  a full-screen editor (e.g. Personal Editor)
- **Speech board interface**
  BASIC subroutines which can be included in a new program. They provide easy access to the parameters computed by the board, as well as control over the board operation, and synchronization of the PC application program with the signal processing routines running on the board.
- **DOS parameters retrieval subroutine**
  Among others, allows for retrieving which language must be used for messages to the user.

**TAILORING POSSIBILITIES**

When this system is set up in a school for the deaf, teachers and speech therapists may use it at various levels, depending on their needs and their computer knowledge:

1. Use the basic application package as-is.
2. Change default run-time parameters (e.g. the pitch range) in control files.
3. Modify existing games, or design new ones, by editing the games description files.
4. Develop new application programs in BASIC.
5. Program new signal-processing routines to run on the board.

Level (4) is obviously where one can take the best advantage of this programmable system, but it is worth noting that levels (2) and (3) also derive from the ease of programming of the system, which enabled a small research team to develop very flexible programs.

**FIRST EXPERIMENTS**

The very first prototypes have been used first in INJS, Paris (1976), then in CEOP (Centre Experimental Orthophonique et Pédagogique, Paris, 1982), but also in Japan and Israel. The programmable system here described has had a much wider distribution during the 1985 spring, as gifts or loans to institutes in France (25 systems) and in many foreign countries: England, Belgium, Germany, Netherlands, Japan, Kuwait, Peru, Saudi Arabia, Spain, Sweden, Switzerland, USA...

The experiment is still of short duration, and some users, for whom it was the first contact with computers, have had some trouble starting with it. But the general reactions are very positive, and show evident the appeal of visual representations to young children, and how they like working with this system.

For sure, it is merely a tool for helping speech training, and the teachers and speech therapists should only use it as a supplemental means within a broader scope of deaf children training. But already, some have started using it extensively. Some of them modified the parameter settings, or even designed and developed new programs.

**CONCLUSION**

The major characteristic of the IBM speech training system is that it is programmable, thus fully "open-ended". It may evolve through simple modifications, or new programs design.

Apart from the direct benefits of this flexibility, it also contributes to the "computer education" of the users, allowing them to better understand their needs and the possibilities and limitations of data processing in their field. This is a key point, because their imagination and creativity will then lead to new breakthroughs in speech training for deaf children.

**REFERENCES**

STRUCTURAL CONSIDERATIONS IN DEVELOPMENT OF A SOFTWARE-BASED REAL-TIME DIGITAL SOUND SPECTROGRAPH MACHINE
L. Robert Morris
Systems and Computer Engineering, Carleton University,
Ottawa, Canada K1S 5B6
and
DSPS Inc, POB 5348, St. F, Ottawa, Canada K2C 3J1

INTRODUCTION
Spectrographic Imaging
A spectrographic display has been essential to the study of sound events for the last four decades. The classic spectrographic technique, in which (about 2 seconds) is first recorded on an annular magnetic surface on the perimeter of a drum. This sound modulates a variable frequency oscillator (VFO) with time. The output of this filter is then input to a bandpass filter. The output of this filter for one pass of the input signal (i.e., for a short duration of the drum), approximates the energy content in the signal at a particular frequency as a function of time. The energy/time function is then wrapped onto the telediagnostic paper wrapped around a conductive drum whose magnetic field (at a fixed frequency) is varied with the drum. The magnetic field is varied with the drum's rotation, so that the VFO's frequency is raised slightly during the drum's rotation. The drum's rotation is then finished and the paper - when the drum's rotation is finished - shows a two-dimensional plot of energy content vs. time. The duration of time for the duration of the sound initially recorded. The mapping shows high spectral energy as black, while the background colour of the paper represents the amount of time available. The entire process takes about 10 minutes.

The characteristic appearance of spectrograms is familiar to speech researchers. Wideband spectrograms clearly show formant structure, and pitch-structure of the original sound. The fine time resolution of these spectrograms reveals the formants and pitch changes in the original signal, which are evidenced by narrow, vertical striations parallel to the frequency axis. The spectrogram can be used for research purposes to separate the individual harmonics of voiced sounds and thus exhibit the horizontal bands whose vertical separation varies with pitch changes. Applications of spectrograms vary but are constrained by significant drawbacks of the mechanical system. The spectrogram production time and the limited dynamic range of the telediagnostic paper. It takes about 10 minutes to see if the correct portion of the sound was captured, which is why Ms. J. correctly uttered in order to demonstrate a certain effect in the final spectrogram.

Ideal Sound Spectrograms
Spectrograms are thus usable. A spectrogram would be able to produce spectrograms in real-time. It would be able to show both wideband and narrowband spectrograms. Pitch-structure of the original sound would be revealed by the spectrogram. The original sound could be labelled, stored, and retrieved at will, depending on the type of analyses. (The original mechanical printing mechanisms with the required speed are prohibitively expensive.) An acceptable strategy would be to have computer images in "real-time" and to be able to request "hard copy" only for cases of particular interest. The spectrogram should be able to continue his real-time work knowing that the hard copy of the requested spectrogram will be provided in a reasonable amount of time, and that its quality will approach (or exceed that of) the classical spectrogram. In practice, there would be a tradeoff between speed, quality, and production, cost of the hard copy device.

Currently, the advent of the fast Fourier transform (1957) and the production of the first "inexpensive" digital computer and associated digital hardware would provide the basis for production of "digital" spectrograms [1]. Briefly, the FFT enabled the essential spectral analysis to be carried out atop 50 times faster than was possible previously. Inexpensive computers meant that such machines could be cheaply produced, and that the associated "digital" processing could be done in real-time, that is fast enough to follow the time evolution of the natural signal. Oppenheim (1970, 1) first demonstrated the concept using a special purpose display. Norris (1970, 3, 4) attempted to combine an off-the-shelf minicomputer and graphics processor in conjunction with specialized programming techniques (4, 5) to demonstrate that near real-time software-based FFT routines were possible. In fact, "soft copy" spectrograms of about a second duration could be produced and displayed in about 5 seconds on an IBM 360/50 minicomputer system. While this cost precluded purchasing such a machine for this application, such machines were occasionally available in other environments, and turning them into a near-real-time spectrograph was a time-saving basis "for free" (6) was an attractive idea.

Technological Basis for a Cost-effective Real-time. Software-based Digital Spectrograph
In 1982, components occurred which, for the first time, made the real-time spectrograph an economically possible machine. First, though general purpose minicomputers were increasing in speed and the single chip microprocessor had replaced the microtread (34). The implementation of simple algorithm turns (STL) present on all such machines and central to the FFT's performance was relatively slow. This was because economical considerations dictated that STL be carried increasing repeated additions via a microwired instruction. Typical execution times were 3.3 to 10 usecs. In the early 1980's, at least one "special purpose" signal processing chip (the DEC 7520) appeared in which a significant portion of the chip's area was devoted to a high speed array multiplier, yielding an STL that was an order of magnitude faster than that of GP micros. However, the price of the chip precluded implementing FFT's on the microtread (34).

In 1985, Texas Instruments introduced the TMS 320-10 DSP microprocessor. This chip, which could provide data memory and could address 4k words of off-chip computer memory, was an extension of the programming techniques utilized earlier [4]. We were able to demonstrate that FFT's of reasonable size (i.e., 256 points) could also be carried out at the required speeds (about 0.15 seconds). A seemingly unrelated event was the appearance of the IBM PC. The adoption of this machine as a "universal" standard led to the proliferation of software "plug-in" function upgrade add-on cards for such machines. Included were the so-called "low-price, low-cost," which required for digital signal processing [34] and which contained a TMS 320 DSP processor with a high cost density and with the capability of widely having the capabilities of the main expensive printers available for the PC market were thought not to be capable of generating the hard copy noted above. Finally, there is the novel Pascal compiler. Turbo Pascal -- compiled the software environment necessary to produce the "real-time sound spectrograph". The real-time and graphics environments of spectrograms did not exist -- even with careful planning -- many iterations were necessary to achieve the desired result. In development of the earlier spectrograph [3], testing of each software required recomposing the "driver" program and then linking the driver module to the software spectrograph. Each change entailed a minimum of about 30 minutes to 10 minutes to 30 minutes to 10 minutes to produce a floppy based machine. Turbo Pascal reduced this time to about 2 minutes on a floppy machine!

SOFTWARE FOR A REAL-TIME DIGITAL SPECTROGRAPH (RSL)
The nuclei of a digital spectrograph are the software modules required to effect the fundamental signal processing operations of sampling, pre-emphasis, windowing, digitization, frequency transform, magnitude extraction, and mapping from linear to logarithmic 4-bit bipolar. We describe the time/space tradeoffs in effecting each operation on the TMS 320 below:

Sampling
The TI-Speech board contains the requisite TMS 320 DSP chip. 4K of program memory, 16K of 16-bit dual port memory (available to both the TI-Speech board and the codec chip). The codec chips at 8 KHz only, through a built in 4 KHz lowpass filter, and smears the analog signal into 6-bit data using the digital algorithm. The frequency signals -- effectively retains about 10 bits of precision. The mixed frequency signals are then processed to a low cost of the TI-Speech board. (A high quality, variable sampling rate 4/360/4000 code) is available on the TI-Speech board!) The 8-bit A/D samples are converted to linear via a table lookup algorithm which reduces the memory space by about 20%. With such applications, it is possible, of course, to gather data from an independent, programmable sample-interval A/D.

Pre-emphasis and Windowing
The spectral representation of voiced speech falls off at about 8 db/octave of increasing frequency. Time-domain pre-emphasis is applied in an attempt to approximately flatten the spectrum so as to lower the dynamic range requirements for display. Windowing of the incoming signal provides spectral smoothing which improves the consistency of spectral
estimates. With \( x(n) \) and \( x(n) \) the input and output samples, both pre-emphasis and windowing are effected by

\[
x(n) = (z(n) - 0.5z(n-1) + w(n), \text{ for } n = 1 \text{ to } N
\]

where \( w(n) \) is the window sample, \( N \), the true window length is a function of the desired effective spectrogram bandwidth and the spectral leakage (cf. [1]). In [2], 256 speech samples are analyzed per spectral cross section. Non-zero \( w(n) \) ranges in length from 64 (wide-band) to 256 (narrow-band) with one's extending the rest of the \( w(n) \) array so as to always result in 128 "interpolated" spectral samples.

**Discrete Fourier Transform**

The well-known FFT is but one of a number of algorithms for efficiently computing the discrete Fourier transform [2]. The speed is increased by the use of the radix-4 FFT. It is possible to effect two N-point real transforms via one N-point complex transform followed by an "unscrambling" phase. The detailed structure of the requisite 256 point complex FFT was described in [7]. Execution time is 6.3 sec., with about 1.5 sec. required for unscrambling; thus 256 point real transforms require about 4.8 sec each.

**Magnitude Extraction and Mapping to a Graphic Display**

The magnitude of each complex spectral sample must be derived. Oftentimes, this involves computing the square root of the sum of the squares of the real and imaginary components. While squaring is efficiently carried out on the TG-2010, square root is not computationally efficient. "approximate" complex-to-magnitude conversion algorithms have been developed and analyzed [3]. Among these, the most accurate of these is well-suited for time-efficient THS 320 implementations. Linear-to-logarithmic conversion is rapidly carried out by Hilitch's method (see S3. Since \( 1.0 \leq r < 4.0 \), a 10-bit linear-to-logarithm amplitude sample can be rapidly converted to the 4-bits required to map to a graphic display.

**Total Time Requirements**

The total computation time per cross section, including all signal processing and plotting, is 8 sec. of which the FFT constitutes half. Thus the importance of high efficiency FFT. Suitable graphics boards for PC's can have a resolution of approximately 840x1000 with 16 levels of intensity. This suggests that two spectrograms, each containing 256 cross sections of 128 spectral samples each, can be plotted side by side with room left for time and frequency calibration. (In the TI 3 plane graphics, resolution is 720x360 with 3 bits per pixel. Combining 2 adjacent pixels to approximate 16 intensity levels yields an effective resolution of 360x720, suitable for vertically stacking two spectrograms of 256 cross sections each.) Since 256 cross sections always require 2.06 sec. to process and display, the spectrogram becomes real-time when the window is shifted 2.06 x 1000/256 = 84 samples per spectral cross section. Higher time resolution -- i.e., a shift < 84 samples -- implies non-realtime processing. However, both spectrograms are always ready within 5 seconds, an insignificant delay.

**Software Organization and Control**

The Turbo Pascal control program (CP) communicates with the TMS 320 DSP software via a serial port handshake using an intrinsic MOVTAR command. "src" and "dst" may be Turbo variables or absolute memory addresses. Although processing and data gathering could overlap, this precludes getting data other than from the A/D. Thus the CP requests that data be put into an 8K word buffer in dual port memory and waits. When the data is ready, CP asks that a cross section be processed. As each cross section is done, CP picks it up, tells the 320 to go on to the next, and plots that cross section. Thus CP CPU plotting and TMS 320 processing are overlapped.

For "real-time" processing, CP begins to be processed and the user speaks at the end of the terminal back. Both spectrograms appear within 6 seconds. If a new time resolution is required, the user types in \( N \). 1 < \( N \leq 64 \), the time axes reallocate (about 2 seconds), and the spectrum repeats if \( N = 92 \). Hard copy of the display spectrogram is produced on an Epson FX 80 printer. After 180/180/185/185 characters (about 5 minutes), no. Other variations of display besides wide and narrow band spectrograms are equal time resolution include dual wide or narrow band displays each of differing time resolution, or display of a spectrogram derived from stored data (e.g., a target) vs. repeated real time input.

**RESULTS AND CONCLUSIONS**

The screen output for the digital spectrograph is shown in Fig. 2. Although photographic half-tone reproduction entails a loss in dynamic range over what is visible on the screen, overall results are comparable to the "negative" of the classic speech spectrogram. The second figure shows a reduced size version of the output on an Epson FX-80 printer. Again, detail is lost since the dots forming this black and white, reduced size image tend to merge. The dynamic range of the system are of course lost in static reproductions. The dual spectrograms' appearance within 5 seconds of speaking, the ability to vary time and frequency resolution at will, and the fact that hard copies need only be requested for suitable screen spectrograms are aspects whose value becomes apparent to the system user. And, the ease and speed with which the control program can be changed to alter the system functionality means that the instrument can adapt to the environment in response to user feedback, something not possible with hard-wired machines.

PC's are now ubiquitous. That about $2000US worth of hardware can, with the required software, turn a PC into a real-time spectrograph with hard copy capability is fortuitous for the speech researcher.

**References**


A MODEL OF THE PERCEPTIVE PHONETICS, ATTENDED BY THE HUMAN MEMORY

C. J. M. Kroch

ul. "Jordan Masaryk" 21 A, 66000 Vambal, Bulgaria

Recent developments in perceptual phonetics, part of the study of human perception of speech are associated with advances in the fields of psycho-linguistics, knowledge engineering and applied AI, pattern recognition, etc.

We assume a neural organization of templates in long-term human memory (LTM) with the following structure:

- there are some primary (atomic) phonetic units expressed by some domain of the space of values of some parameters (e.g., in corresponding to a single measurable physical characteristic) the rest of the units being compound and corresponding to composite systems of domains of the parameters;
- to compound templates there correspond parameters of two types (type A and type B).

The units of type A are compound units for which the set of characteristics is the same for any two opposite units and these are distinguished only by the internal values of the compound parameters, e.g., the compound characteristics of accented and unaccented syllables; they both have the same set of parameters, such as duration of the vocal part, duration of the consonant part, intensity of the syllable peak, frequency of the basic tone, etc.; they have non-intersecting regions of values of the compound parameters (so that syllables with or without accent could be told apart).

The units of type B are distinguished from opposite units by the existence of a parameter which is absent in the representation of the counterpart (any one of which is an example of unit A);
- thus, in contrast to units of type A, the identification of units of type B may be based on a specific set of characteristics and not on an internal compound characteristics (as the units of type A). Based on experimental data, a hypothesis is put forward in that the compound parameters of units of type B can themselves be composed by units of type A, in particular, distinct differential characteristics of the phonemes occurring in different units, can be established by summing up the values of its components.

- templates in LTM of the phonetic units which correspond to parameters in the space of values of which any two realizations are indistinguishable within the limits of the region leads to perceptually indistinguishable realizations. Such a view on the functionality of the templates is founded on ignoring in the perception of the basis of the language of the variations which are small. On the other hand, identical reaction to physical features that are "near" enough is histologically natural. In this respect it resembles the law of Tall or nothing.

- an immediate neighbourhood of a Zip is the zone of similarity to the template (ZIP).

The ZIPs of distinct phonetic units do not intersect, moreover they have non-intersecting closures in the topology, generated by the notion of nearness, while the STM may well have non-empty common parts and this is one of the explanations of ambiguous perception.

For units that do not have a corresponding sound image the existence of a zone of identical reactions can also be conjectured as well as of zones of similarity;
- the categorical character of speech sound perception is not, what we do proposed the notion of different speech sounds being comprehended in two completely different ways: "categorical" and "non-categorical";
- the boundaries of the zones (in particular, of STM) are carried out with the templates of some phonemes, but with templates of some combinations. If in the set of templates in the perceptual basis of the human mind there is no suitable template (exactly fitting) the result is to be the nearest such templates (in the topology) and to all combinations of them until a suitable combination is found and a satisfactory similarity is established. Of course, another possible explanation is that phonetic templates are indeed the templates of primary units and the perception of a sound word a simultaneous correction is taking place. But data from 2 and 5 supports the view that this is not the case and that the units of the primary perception are not the phonemes, but certain their compounds, in particular, the syllables. One more reason for this is the fact that in experiments with perception of syllables the reaction time for single phonemes is much greater than the reaction time for syllables themselves, thus one is bound to insist that the real formative units are the syllables.

On Fig. 2 the same is in accordance with data about perceptual phonetics (PP) some of which was discussed above a flowchart of the hypothetical phonetic verbal perception, based on the system of human memory. In conclusion, we mention some work done on the topology in the space of perception parameters: the notion of nearness must adequately serve to generate a topology which is not a Hausdorff one, but what we can claim at
most is that this topology is a $T_1$-topology.

It is a matter of further investigation to
decide whether some kind of proximity space
would not fit the picture better.
VOEwEL+CONSONANT TIMING ACROSS SPEAKERS.

C. BEMDMT & L. ABAY

*Centre National d'Etudes des Télécommunications, Issy-les-Moulineaux
**Institut de la Communication Parlée, Grenoble

INTRODUCTION

The search for invariants in speech literature has received these last years new support from the so-called Action Theory [1].

The aim of this paper is to examine inter-speaker variations and to find out whether or not acoustic relative timing is preserved as it has been reported for rate and stress changes in E.M.G. [2] and for movement [3].

1. GUBING EVENTS

When an American lobster walks [4], the step cycle of each leg can easily be detected, like in vertebrate and human locomotion [5].

So, a PERIOD is a cycle - can be defined as the interval between two successive occurrences of the same type of event. These events can be plotted from motoneuron activity, E.M.G. signals, as well as movement tracings. They can be defined as onsets, offsets, peaks amplitude, etc., in any of these channels.

Sometimes, each channel could be analysed as a combination of several subchannels: this is apparently the method used in the search of cycles in speech analysis.

In their pilot study, TULLER & al. [2] used integrated E.M.G. onsets of muscle activity relevant to lip closing (orbicularis for [p]), tongue fronting (genioglossus for [i]), and jaw lowering (anterior belly of diastragic for [a]) in [papip] utterances. Attributing differences between vowels and consonants, they were able to define a VOCALIC CYCLE [a-i] and a CONSONANT LATENCY, i.e. the delay in the onset of consonant activity within the vowel-to-vowel period.

They noted [2] that such a practice reflects a conception of speech production as a continuous vowel-to-vowel cycle within which consonantal activity is superimposed [7]. They agreed with the fact that such an analysis is quite different from that used in locomotion, where the timing of extension-flexion "is not described as continuous flexion-to-flexion with extension superimposed on this basic cycle" [7]. We came across this problem in an earlier study [3] where they plotted events from the kinematics of the jaw (one channel) and those of the lips (two channels) for the cycles of the jaw in [babab] utterances which was associated with the vocalic cycle [a-i]. In fact, the onset of jaw lowering for [a] is synchronous with the onset of the lower lip lowering for [b] release. An interdependency remains even when jaw movement has been subtracted from global lip movement, and separating the channels is thus difficult (at best, upper lip cycle for consonants vs. jaw and lower lip for vowels and consonants).

The partitioning of tongue temporal function into two channels for [kakak] utterances, in a recent paper by MUNHALL [6], reiterates the question.

Intervals between onsets of tongue lowering for vowels determine the periods, and within-period onsets of tongue raising for consonants separate the latencies: "These intervals are the within-articulator counterparts to the period and latency measures used by TULLER & al." [6].

We too analysed the speech audio signal into channels, trying to localize acoustic discontinuities (events) as targets of articulatory (glottal and supraglottal) events [7].

For supraglottal events, we used evidences for closure-striction (C) and release (R) in GvVv; Gv; lobes with plosives and fricatives [8] as follows:

- C: "The point where high frequency components of the periodic wave disappear" [2].
- R: The point where the relatively high aperiodic energy of the burst (glission and/or friction) yields to the low energy noise of aspiration (or murm). i.e. the end of the "initial phase" [9] for plosives and fricatives.

R reveals the closure (or striction) achievement (see [3], Fig. 1 for lip closing).

Our method differs from the one used by TULLER & al. [2] where acoustic events are marked on CV syllables, from closure to closure. To our knowledge, they did not analyze the acoustic signal in terms of period-latency along R-to-B vocalic cycles relating them to the articulatory ones.

The acoustic periods simply appear to be delayed in comparison with the articulatory ones (cf. [3], Fig. 1; [6], Fig. 1).

Apparently, this relationship had not been noticed until recently [11]; thus omitting when calculating latency/period ratios (L/P) well known measurements (V/W+C) in the suprasegmental literature [12], [13].

2. RELATIONS BETWEEN LATENCY AND PERIOD

In search of relative invariance in E.M.G. averaged repetitions, TULLER & al. [2] obtained constant L/P ratios and correlations. Later, TULLER & KELSO [3] used movement tracings to obtain correlations, taking into account different repetitions. The high correlations observed were obtained across different conditions of stress and rate. More recently, LOPFIVIST [11] and MUNHALL [6] found converging results for over-all conditions, but somewhat lower for within condition correlations. We transposed this analysis to our acoustic data and obtained the same tendencies. Each speaker exhibited low correlations but an overall analysis based on the four speakers highly increased the correlation results [14] (Fig. 2).

Does this mean that phase relationships are preserved in articulatory and acoustic cycles - especially in the well known VC domain - under transformations in rate, stress and across speakers?

3. A PART-WHOLE PROBLEM

Since BARK's objections [15] to the statistical "artifact" emerging from high correlations between self-included intervals (latency as a part of the period) we submitted the problem for publication and suggested similar conclusions as exposed by MUNHALL [6]. We can proceed to further observations. In the case of a whole constituted by two overlapping parts: any part-whole correlation is always higher than the correlation between the two parts; the higher the part-whole correlation, the higher the variability of this part; there is a quite simple relation between the vowel to cycle part-whole correlation and the standard deviations of the parts, as shown in Fig. 3.

Consequently the between-parts correlation value must be available in any part-whole correlation study. It would henceforth be easier to argue on part-to-part scattergrams.

For example, coming back to TULLER & KELSO diagram [3], we replotted their data and subtracted the latency from the period.
Looking out these new scattergrams e.g. for [babab] (Fig. 4), it is obvious that:

Firstly, intraclass correlations are low. For such a results, see PORT & al. [16], whose viewpoint considers only tendencies in means values as being relevant.

Secondly, mean trends differ under stressed (3-4) and unstressed (1-2) conditions, with a change in intonation.

Vowel and consonant vary in the same way, only if unstressed. We can thus talk of some phase relationship conservation in this last condition. When the syllable is stressed, we observe the well known fact that consonant prevails mainly on the vowel across changes in rate.

Looking at our inter-speaker variations, we can see in Fig. 5 that vowels and consonants follow the same mean tendencies, both increasing from the fastest to the slowest speaker in our pre-stressed condition.

4. Back to ration

The study of ratios (L/P, V/V+L+C) is the best means of examining a possible preservation of phase relationship. So E.M.G. data constancy [2] must be linked to movement findings [3] and also to acoustic measurements.

As concerns movement, replots from [3][see Fig. 6]: same [babab] as in Fig. 4 display a clear tendency for the vowel to undergo the greater part of the compression/expansion effects, whatever the conditions are (the change being mainly on the vowel or the consonant).

For acoustic data and for similar paradigms (quantity and rate [14]; quantity and sentence accent [13]; etc.) we observe the same behavior. More generally, strong segments ("long" vowels or consonants) account for the greater part of the changes [12].

Fig. 7 shows, for our acoustic measurements, the same trend for the vowel to increase proportionally from the fastest to the slowest speaker.

REFERENCES


CONSONANT-TO-VOWEL DURATIONAL EFFECTS IN ITALIAN

Pier Luigi SALZA and Stefano SANDRI

CSELT - Centro Studi e Laboratori Telecomunicazioni
Via G. Reiss Romoli 274 - 10148 TORINO (Italy)

INTRODUCTION

Results published in the literature on the subject can be roughly shared into three categories, according to different hypotheses.

1) Phonetic context effects for many languages could be related to articulatory constraints: in connected English, for instance, stressed vowel duration appears strongly affected by the voicing of the following Plosive, Plositive or Nasal, mainly in prenasal syllable [1]. Early measures carried out on isolated Italian words showed similar, but weaker, trends [3]. On the other hand, voicing effect does not seem to be an universal phonetic feature [3].

2) Further investigations on isolated words demonstrated that Italian stressed vowel duration is also affected by syllable structure (open vs closed): a sort of compensation would exist between the duration of the vowel and that of the following consonant [4]. Timing effects could be viewed as a part of phonology of a language, where the global VC duration could be considered as an invariant prosodic macrounit (anticipatory trans-syllabic effect), according e.g. to the results of [5], based on frame sentences. As far as English is concerned similar data can be found in [6], [7].

3) Another research, based on both acoustic and perceptual measures and using again frame sentences, stated that stressed as well as unstressed vowel duration has little or no phonological value in Italian, but rather for each intervocalic consonant a critical threshold of C to preceding V duration ratio could be precisely defined [8].

Of course, different results might be partially due to different corpus selections and different experimental procedures. In this investigation we try to assess the above hypotheses on a wide corpus of polysyllabic words, as described below.

THE SPEECH MATERIAL

The corpus was previously designed for a comprehensive set of microprosodic measures [9], where the nonsense paradigm PACAP'A was chosen as speech material. The test syllables are the unstressed ones, as this condition allows us to avoid cross correlation between phonetic context and stress effects. Thus, for example, in the word /pastap/ the measured segments are /asta/, that is the intervocalic consonant context and both bordering /a/’s. A professional speaker uttered the words five times in the sentence frame “Devo dire... claramente”. According to the exposed hypotheses, consonant contexts must be considered separately whether they occur in open or in closed syllable.

An open syllable is defined by one of the following conditions: PA/PAPA, PA/XLAPA’; where: C: any single consonant; X: Plosive or Fricative; L: Liquid; S: syllable boundary. The open syllable corpus is reported in Tab. 1, where consonants are classified by voicing and manner of articulation; place features are expressed in [9]. Single consonant occurrences refer to northern variety of Italian.

Any different consonant context is meant to occur in closed syllable, including the SC clusters [10]. Corpus of closed syllables is shown in Tab. 2.

RESULTS

Fig. 1 collects measured durations of unstressed /a/ preceding and following the various consonantal contexts indicated on the horizontal axis: A) open, B) closed syllable. Average individual durational values were pooled on numerical basis within a range of ±7 ms and plotted in the figure according to decreasing order of the preceding vowel duration, for voiced and unvoiced cases separately.

Articulatory constraints produce considerable effects, in agreement with hypothesis 1): voiced consonants lengthen both preceding and following /a/ with respect to voiceless ones, by a factor depending on articulatory class. Clusters tend to behave like their first member alone. "Long" palatals do not demonstrate to affect vowels differently from geminates. In open syllable the voicing lengthening factors of the first vowel are for each class: F 117, A 25%, P 47%; while in closed syllable: S 39%, F 39%, A 32%, P 29%. Nasal and Liquid lengthen the preceding vowel by 23% in open and 10% in closed syllable, with respect to unvoiced Plosive.

Since differences between voiced and unvoiced cases are maintained in open syllable, the presence of a syllable boundary does not seem to seriously influence consonant effects on vowel duration [6]. Average absolute values are very near in the two cases, compared for the same articulatory classes. By averaging among all classes, excluded S in open syllable that has not unvoiced counterpart, the following average values are obtained: in open case 117 ms, 93 ms, in closed case 110 ms, 94 ms, for voiced and unvoiced contexts respectively. Shorter values are observed in N, L classes of closed syllable, and in rightward extended clusters.

Moreover, it comes out clearly the general trend of the following /a/ to be always shorter than the preceding one by a factor mostly independent of voicing, syllable structure and absolute durations. On the average the second /a/ is the 88% of the first one in open syllable, and the 83% in closed one. Thus, it is likely to reflect a rhythmic constraint, according to which the unstressed nucleus in medial pre-stressed position of a polysyllable undergoes greater compression [11].
Compensation exists between vowel and following consonant durations with respect to articulatory closure and voicing of consonants, in open and closed syllables distinctively. But it does not exist between open and closed conditions, as shown by the following average VC values: open syllable 191 ms, 191 ms, closed syllable 241 ms, 259 ms, for voiced and unvoiced contexts respectively.

It appears that the noticeable difference between average VC durations in the two conditions is due to consonant. In few cases it is due to individual behaviours of consonants in clusters, e.g. Nasal and liquid show opposite modification trends with respect to their inherent durations [9]. But most differences are realized by gemination of consonants, almost independently on their articulatory class and on the presence of a further following consonant. In fact, by comparing open and closed cases, almost any context reaches average VC durations comprised between 100 and 210 ms, except geminates and geminated plus liquid cases, whose average VC durations are from 250 through 300 ms; this sharp magnitude change differentiates them clearly. Hence, the discriminant issue is whether the following consonant is single or geminated, namely the phonological length of consonant.

If, at the moment, the hypothesis 2) about duration compensation between consonant and vowel in tautosylabic and heterosylabic case has to be rejected, significant evidence is gained that supports the hypothesis 3) concerning the existence of critical thresholds of C/V duration ratios [8]: ratios ranging from 0.5 to 1.25 are shared by both open and closed cases concerning short consonants, while geminates, either followed by another consonant or not, present C/V ratios comprised between 1.8 and 2.25, with the only exception of /r/ and /d/.

CONCLUSIONS

The factors here considered in relation with open/closed syllable contrast, namely articulatory constraints, consonant to preceding vowel duration compensation, and C/V duration ratio, were assessed to be quite significative independent variables. These results suggest that above drawn hypotheses are not necessarily incompatible.

ACKNOWLEDGMENTS

This work was partially supported by an EEC contract in ESPRIT Project No. 64. The authors wish to thank E. Foti for his appreciable assistance in speech database processing.

REFERENCES

AN ACOUSTIC STUDY OF VOWEL LENGTHENING AND PAUSING IN SYNTACTIC TIMING

P. Avery, W. Cichocki, W.J. Idsardi, K. Rice, J. Archibald

Department of Linguistics, University of Toronto and
Department of French, University of New Brunswick
(Cichocki)

INTRODUCTION

The phonetic implementation rules of vowel lengthening (VL) and pausing are considered to be a phonetic encoding of hierarchical syntactic structure. Research has shown that vowels are lengthened at major constituent breaks and that these constituent breaks also correspond to potential pause breaks in the speech stream (Klatt, 1975; Grosjean, Grosjean and Lane, 1979; Cooper and Paccia-Cooper, 1980). While VL and pausing have the same environment, studies of rhythmic structure have generally avoided unifying these phenomena (cf. Liberman, 1978 and Duez, 1982).

In this study, we subject to empirical testing a theory of syntactic timing found in Selkirk (1984). This theory is able to unify VL and pausing - showing that they have the same environment because they are part of the same phenomenon. Selkirk proposes a model of grammar in which surface syntactic structure is mapped onto an autonomous phonological structure represented in terms of a metrical grid (see Liberman and Prince, 1977; Prince, 1983). In mapping from syntactic structure to phonological structure, silent grid positions (demiboots) are added by the following procedure, which is a phonological interpretation of syntactic structure:

(1) Silent Demiboot Addition (SDA) (Selkirk, 1984: p. 314)
Add a silent demiboot at the end (right extreme) of the melodic grid aligned with:

a. a word,
b. a word that is the head of a non-adjunct constituent,
c. a phrase,
d. a daughter phrase of $S$.

The phonological representation of the utterance shown in (2) illustrates how the rules of SDA are applied.

(2) x
    x
    x
    x
    x
    x
    x
    x
    x
    x
    x

[fill]

6 6 6 6 6

the boat in the bay sank

Each syllable (6) in the utterance is associated with a grid position (x), an abstract timing unit, where each grid mark occupies approximately the same amount of time. The relative prominence of a given syllable is determined by the height of the grid column, while junctural properties are determined by the rule governed insertion of silent grid positions. In (2), these are the x's to the right of 'boat', 'bay' and 'sank'. Selkirk's claim is that the silent grid positions between words can remain unfilled (pause) or be filled by spreading of the preceding syllable into these positions (length), as indicated by the slanted lines in (2). For example, five silent grid positions separate 'bay' and 'sank' in (2). Spreading of the syllable could fill all or only some of the positions. Those which are filled by spreading will be realized as length on the vowel and those which are unfilled will be realized as pause. Selkirk suggests that the degree to which a particular vowel can be lengthened is governed by the 'stretchability' of the syllable.

Given this linguistic representation, two clear predictions emerge. First, the total time of the vowel plus the pause should increase as the number of silent grid positions as added by SDA increases. Second, an inverse relation should exist between VL and pausing. This is because spreading must be seen as an optional rule and, if it does not apply, then the vowel should not be lengthened. However, the grid positions are still present and should be realized as pause. In the preliminary study we report on in this paper, our main interest is in the first of these predictions.

THE STUDY

Materials

Ten sentences were constructed with subject NPs of increasing complexity and thus more grid marks separating the subject NP from the initial element of the VP. (See (1) for how grid marks are inserted.) All subject NPs were completed by the word 'bay' and the VPs began with the word 'seems'. These words were chosen to facilitate acoustic analysis as the final portion of the vocalic element is clearly separated from the beginning of the initial consonant of 'seems'.

The sentences were not contextualized but the subjects were asked to place the main stress of the sentence on the word 'seems', that is, away from the word 'bay'. This ensures that the length of 'bay' cannot be attributed to main stress or focus factors. The ten target sentences were randomized among ten distractor sentences. Subjects familiarized themselves with the material before the recording. They were instructed to read each sentence at a 'normal' rate; their productions were recorded on a Uher Report 400 tape recorder. Some examples of the sentences are given in (3) to (5) below.

(3) The bay seems to be calm today.
(4) The house on the shore of the bay seems...
(5) The man you say you sailing on the bay...

The number of silent demi-beats ranged from 3 to 10. The only missing data point was at 9 silent demi-beats, two sentences had 3, two had 5 and two had 6.

Subjects

Subjects were seven native speakers of English who are students at the University of Toronto. Both males and females participated. Each served individually in sessions lasting approximately fifteen minutes.

Analysis

Wide-band spectrograms (300 Hz and 600 Hz bandwidth) were made with an RT-100 real-time colour spectograph (developed by P. Martin of the University of Toronto). Vowel and pause durations of the target environment were measured (in msec) using conventional segmental criteria.

Initial analysis of the measurement data involved graphing individual subject's durations (pause, VL, and pause plus VL) against number of grid marks. Results were then analyzed using a multiple regression model. In this paper we report only on correlational data as the sample size was rather small and we feel that interpretation of the regression equations with limited data is not feasible.
Results
Preliminary examination of the graphs revealed that the overall pattern of the data was as the theory had predicted. However, two sentences behaved in a consistent yet unexpected manner (VL plus pausing was exceptionally long). One involved the NP 'the larger bay' and the other a NP with a double PP modifier, 'the house with the awnings near the bay.' These sentences were excluded from the analysis as we felt that the subjects had difficulty in placing the focus on the VP rather than the subject NP. While this decision certainly has consequences for the final analysis, we did not feel that the exclusion was unwarranted. In the first place, it was our intention to control for any effects of focus and main stress, as those factors might confound the results. Secondly, it is not clear to us that the assignment of silent grid positions is correct in these cases particularly in the case of double PP modification. The syntactic structure in question is left-branching and because of this does not receive as many grid marks as a uniformly right-branching structure. This again appears to be a problem for the rules of SDA, suggesting that some revision may be necessary.

The results of our analysis of the remaining eight target sentences are summarized in the correlation matrices in (6). For all subjects, there is a strong association between the number of grid marks and the overall duration of vowel length plus pause. This verifies one of the main predictions of the syntactic timing theory.

(6) Subject  Grid/Vowel Grid/Total Grid/Pause Pause/Vowel
1  .97*** .94*** x x
2  .62** .62** x x
3  .79* .85** x x
4  .51 .63** .86** .61
5  .65 .76* .73* .56
6  .63 .69 .73* .65
7  .24 .69 .76* .20
*** p < .001, ** p < .01, * p < .05 xx, no pausing

Several subjects (4-7) show a strong relationship between the number of silent demi-beats and the length of the pause (Grid/Pause). These same subjects do not display a strong relationship between the length of the vowel and the number of grid marks (Grid/Vowel). This indicates that there are two basic patterns in the data, one for those who pause and one for those who make minimal use of pausing. When pausing is employed the strongest relationship holds between the grid and the pause time. Note, however, number of grid marks and the total of the vowel plus pause time (Grid/Total) is almost identical. This is because there is very little pausing in the data of these three subjects.

Discussion
The results of the present experiment support the theoretical model developed in Selkirk (1984) but suggest that some modifications may be necessary. It is apparent that some subjects did not employ pausing at all. These subjects were, however, displaying a pattern of vowel lengthening which could be handled very easily by the spreading theory. Those subjects who were pausing did so for a much longer time than would be expected if vowel lengthening and pausing are exactly the same phenomenon. We feel that the results suggest that both VL and pausing may be instantiations of silent grid positions but that the timing phenomenon must be treated separately. In this way, a pause which occupies a certain number of grid positions would not necessarily have the same timing as a vowel which is spread into the same number of positions. Selkirk states that syllables may be variable with respect to their 'stretchability' and we conjecture that pauses are more 'stretchable' than syllables. This means that the phonetic implementation rules may treat pause and VL in a different fashion with respect to timing.

CONCLUSION
One issue not addressed in this study is the effect of the vertical grid column on syntactic timing. The results presented are supportive of the model of phonological representation of syntactic timing outlined in the introduction. In particular, the overall duration of VL and pause is closely related to the horizontal grid marks inserted by Silent Demibeat Addition. Stress also affects vowel duration in English but grid theory represents stress only vertically. The mapping rules which assign time to grid marks must be modified to take the durational effect of stress into account. We are currently engaged in further research in which the target word is contained in a greater variety of syntactic contexts and where we attempt to uncover the effect of stress and speech rate on duration. We are also studying the predicted inverse relationship between VL and pausing.

REFERENCES
ACOUSTIC MEASURES OF RHYTHM IN INFANTS' BABBLING, OR "ALL GOD'S CHILDREN GOT RHYTHM"

C. Bickley, B. Lindblom, L. Roug

Rm 36-521, Department of Electrical Engineering and Computer Science and Research Laboratory of Electronics, M. I. T., Cambridge, MA, USA 02139; Institutionen för Lingvistik, Stockholms Universitet, 106 31 Stockholm, Sweden

Investigators of infant motor behavior have observed a general tendency for movements to form repetitive patterns (Thelen, 1981). Rhythmic sequences of consonant-vowel units appear abruptly in children's vocalizations at six to eight months of age and represent a major milestone in descriptions of a child's vocal development. A remarkable characteristic of this style of vocalization, which is known as canonical babbling, is its speech-likeness and quasi-syllabic organization. Figure 1 compares spectrograms and waveform envelopes of a child's utterances at two stages of development: an example of canonical babbling (left) and a later production of a meaningful word (right). Note the similar sequences of articulatory closures and openings that the two spectrograms suggest.

Figure 1: Spectrograms and waveform envelopes of a canonical babble (left) produced by a child of 45 weeks of age and of a meaningful word (right) produced by the same child at two and a half years.

Gillner (1986) lists several characteristics of canonical babbling. Two of these can be interpreted as descriptors of rhythm: (1) amplitude differences between peaks and valleys in the power envelope of at least 10 dB; and (2) peak-to-peak durations of 100 to 500 ms.

Holmgren et al. (1986) bring up the possibility that children's vocalizations might show evidence of the temporal and rhythmic features of canonical babbling independently of other characteristics. The present paper is a follow-up of Holmgren et al.'s suggestion in that we shall propose and evaluate two methods for investigating certain rhythm-related temporal properties of pre-speech vocalizations from acoustic recordings. We will attempt to answer the following questions: When do sequences of quasi-syllabic units first appear in an infant's vocalization? What temporal regularities do they exhibit?

OVERVIEW

One way of studying the rhythm and timing of speechlike phenomena in infant vocalization might be to measure the duration of intervals corresponding to "syllables," "vowels" and "consonants." A major correlate of a syllable in adult speech can often be found in the envelope of the speech waveform which will typically represent consonant segments as low-amplitude, relatively short intervals and vowels as portions of higher amplitude and somewhat longer duration. Segmentation into such units is known to be difficult for adult speech and would be expected to be no less so for pre-speech utterances. Therefore, we have chosen a method which attempts to establish the presence or absence of rhythmic regularities on the basis of low-frequency periodicities in the waveform envelope. These periodicities which occur at a "syllable" rate reflect in most cases articulatory movements such as an alternation of open-close gestures. Similar patterns are occasionally produced by children by modulating the amplitude of the voice source while maintaining a relatively fixed vocal-tract configuration.

METHOD

Each waveform envelope was smoothed by low-pass filtering at 20 Hz. For each smoothed envelope, the autocorrelation function (AC) and discrete Fourier transform (DFT) were calculated and were analyzed for evidence of syllable-like intervals of high and low amplitude. A peak in the AC function at a non-zero time identifies the duration of the interval between two such intervals. The lowest frequency peak in the DFT spectrum corresponds to the length of the utterance. A peak of greater amplitude at a higher frequency would correspond to the syllable rate of the utterance. If no such peak exists, then the utterance envelope contains no syllable-like alternation of intervals of high and low amplitude. AC functions and DFT spectra exhibiting these patterns would be produced by sequences of relatively open, vowel-like intervals alternating with consonant-like closures. Figure 2 shows spectrograms, waveform envelopes, AC functions and DFT spectra for an utterance which is classified as rhythmic (left) and one which is not (right).

Figure 2: Spectrograms, waveform envelopes, AC functions and DFT spectra of rhythmic (left) and non-rhythmic (right) utterances. These utterances were produced by a child of age 45 weeks.
Data comparing "syllable" rates as determined from hand measurements of vowel-onset intervals to the rates calculated by the AC and DFT methods for one child's utterances are shown in Fig. 3. These data were compiled from measurements of utterances which consist of clear intervals of high and low amplitude. In most of the cases, the measured and calculated vowel-onset rates are similar. For utterances which consist of less sharply differentiated intervals of high and low amplitude, and therefore for which the calculation of an onset rate is less obvious, the automatic methods often indicate no periodicity, and thus no rhythmic variation in the waveform envelope.

![Figure 3: Correspondence between measured and calculated vowel-onset rates. The abscissa represents the duration between vowel onsets as measured by hand from waveforms; the ordinate, the vowel-onset rates calculated by the AC (circles) and DFT (triangles) methods. Perfect matches between measured and calculated values are represented by points on the dashed line.](image)

RESULTS

The results of the AC and DFT analyses show a relation to perceptual judgments of rhythm in pre-speech utterances. Three listeners judged utterances on a subjective scale ranging from "definitely rhythmic" to "no rhythmic." These judgments and the results of the automatic classification are shown for one child in Fig. 4. The percent of utterances exhibiting rhythm is shown as a function of the age of the child. The data show a correspondence between the automatic classifications and perceptual judgments.

![Figure 4: Percentages of rhythmic utterances as judged by listening and an classified automatically. The open circles represent the percentage of utterances classified rhythmic by either the AC or DFT method; the filled circles the percentage judged "definitely rhythmic."](image)

The onset of rhythm in pre-speech vocalizations can be inferred from data such as those shown in the last figure. At weeks 21 and 25, few utterances (less than 30 percent) were found to contain any evidence of rhythmic variation in the waveform envelope. At week 31, there is an abrupt increase in the percent of utterances classified as rhythmic.

The onset of canonical babbling for this child was determined from a transcription study to be at week 33 (Holmberg et al., 1986). We conclude from these data that the onset of rhythm precedes the onset of canonical babbling by at least two weeks for this child. Figure 5 shows the percent of rhythmic utterances (as classified by the automatic methods) overlaid on a graph of the percent of canonical babbles as determined from the transcription study at seven points in time. We have examined data for a number of other children in this same manner and conclude that the general trend is for the onset of rhythm to precede the onset of canonical babbling.

![Figure 5: Percentages of canonical babbles as a function of age as determined from transcriptions (filled circles) and of utterances classified as rhythmic by the automatic procedure (open circles).](image)

CONCLUSIONS

An acoustic approach to investigating the appearance of "protosyllables" in infant vocalizations has been proposed. Procedures for the automatic measurement of the temporal organization of syllable-like events have been outlined. These methods examine the waveform envelope for the absence or presence of low-frequency periodicities related primarily to opening-closing gestures of the child's vocal tract. The methods were found to classify pre-speech utterances as rhythmic or non-rhythmic in a manner closely mirroring perceptual judgments by human listeners. The procedure can be used to provide quantitative information documenting developmental milestones (e.g., canonical babbling in Uller's metaphorical framework). Among other advantages is the automatic nature of this procedure; thus it can be applied more easily than transcription techniques, to large bodies of data. This technique could be used as one component of a diagnostic evaluation of a young child's verbal development.

This work was supported in part by a Whitaker Health Science Fellowship, the First of May Flower Annual Campaign for Children's Health and the Center for Advanced Study in the Behavioral Sciences.

REFERENCES


LA VOIX DE L'ENFANT A LA PERIODE DU BABELLAGE (8-12 mois)

G. Konopczyński
Laboratoire de Phonétique, Université de Besançon (France)

Si les études sur la voix de l'adulte connaissent un vif intérêt, notamment en raison des recherches sur la reconnaissance automatique du locuteur, les travaux sur la voix de l'enfant et son évolution sont peu nombreux. Seules deux tranches d'âge ont retenu l'attention des chercheurs : la période néo-natale où les paramètres du cri ont été étudiés par des équipes médicales, en raison des indications qu'ils apportent pour le dépistage de certaines pathologies (Rosenhouse 1980, Kaskinen 1982) et, à l'autre extrémité, la période de puberté qui a également donné lieu à un grand nombre de travaux (Holm 1971, Hollien 1965, 1972, 1977, 1980, 1983). En revanche, l'évolution de la voix aux âges intermédiaires, notamment durant la période de babillage a été un sujet pratiquement ignoré. Certes, la littérature apporte quelques renseignements, sur la hauteur de la voix enfantine (bilan in Konopczyński 1983). Mais la plupart des travaux posent de réels problèmes méthodologiques (Haus 1979). En effet, l'analyse porte généralement sur des corpus hétérogènes couvrant plusieurs années, mêlant cris et vocalisations, sans tenir compte de leur contexte d'émission. La seule inclusion du cri dans le corpus ne suffit pas à se faire une idée précise de la voix de l'enfant dans le médium inférieur, qui est un enfant de trois à quatorze ans. Pour cette raison, nous avons choisi de ne pas nous limiter à l'étude de la voix de l'enfant dans le médium de production (Jasis) et celles produites en interaction avec un adulte (Proto-Language). Chacune de ces catégories d'énoncés possède une structure syllabique, temporelle (Konopczyński 1983, 1985) et articulatoire qui sont propres. Notons toutefois que l'étude ne se limite pas aux aspects de la voix produite par les enfants, mais aussi à leur usage dans le contexte de la production de la parole (Jasis) et des échanges interlocutoires (Proto-Language).

L'étude portera sur 2230 énoncés émis par trois sujets (2M, 1F). Ils sont analysés à l'aide du décodeur de méloïdie P.M. (Martin & al. 1985) dont les résultats sont complétés par des analyses sonographiques.

Pour étudier la voix et son évolution, nous avons été amenés à définir un fondamental usuel (Po-u) qui représente la dynamique de base du locuteur, c'est-à-dire la hauteur à laquelle la voix se place naturellement, mécaniquement, sans aucune intention particulière ne soit réalisée. Ce Po-u nous servira de référence. Nous nous interrogerons sur le registre vocal des émissions solitaires du Jasis, et celui du Proto-language des interactions avec un adulte.

1. LE FONDAMENTAL USUEL (Po-u)

Il se repère très aisément chez l'enfant. Celui-ci émet en effet un nombre important de productions vocales quasi inconscientes et neutres, qui ont de véritables en-ouvertures, brefs (M=220 ms, ET=65%, ECT=615 ms) monotonés, d'intensité faible. Ils sont totalement différents des vocalisations fluctuantes du jeu vocal ou des émissions de PL. Malgré leur fréquence d'occurrence élevée, ces [a] restent généralement ignorés en raison de leur vacuité. Ils présentent pourtant pour le moins étonnante à ce stade d'évolution langagière : en effet, alors que l'instabilité des émissions, à tous les niveaux, règne en maître, ce Po-u affiche une remarquable stabilité, comme le montrent les documents qui démontrent l'importance des sujets féminins. Pour les deux autres sujets, les résultats sont comparables:

Po-u: 57Hz E.T. 82Hz tessiture: 320-460Hz 53% (M) 325Hz 77 220-470 Hz 9% (F)

L'analyse des histogrammes et des courbes cumulatives, dont le document joint est un exemple, fait ressortir le peu de dispersion du Po-u. La zone tonale la plus employée est celle comprise entre 300 et 400 Hz. Les courbes présentent une pente absolument constante et montrent que 75% de la tessiture se situe entre 260 et 400 Hz. Les zones en deçà de 240 ou au-delà de 460 Hz ne sont pratiquement jamais utilisées pour ce type d'émissions, dont aucune ne dépasse 500 Hz.

Outre la constance du Po-u dans la zone tonale, on notera que l'enfant domine de mieux en mieux sa voix, puisque les écart-types, relativement élevés aux mois 8 et 9, se réduisent à moitié au mois 10, la tessiture, d'abord comprise à 75% entre 260 et 400 Hz, soit sept octaves, se rétrécit pour être à dix mois comprise à 95% entre 270 et 380 Hz, aucun [a] ne dépassant plus les 420 Hz.

Soulignons également que l'intensité des [a] est faible, et surtout qu'elle ne change pas lorsque le niveau sonore environnant augmente. Ceci montre que l'enfant est dans un état neutre, qu'il ne cherche ni à attirer l'attention, ni à communiquer. Ce paramètre n'évolue pas non plus avec l'âge ; l'intensité moyenne ne dépasse jamais 22 dB et l'écart-type est toujours inférieur à 6 dB.

2. LES VOCALISATIONS SOLITAIRES DU JASIS.

Le document d'accompagnement montre que le Po-moyen se distingue du Po-u par son élevation, sa dispersion et son instabilité. Les histogrammes et courbes cumulatives révèlent clairement ce fait : aucune gamme de fréquence n'est majoritaire, les variations de voix se répartissent inégalement sur toute l'échelle comprise entre 200 et 500 Hz (cf. pente douce des courbes) et 30 à 40% des énoncés dépassent 500 Hz durant tout le mois 9. Enfin, alors que le Po-u atteint sa stabilité dès le début de ce mois il y a évolution dans le Po-m qui baisse progressivement ; la dispersion diminue également, tout en restant bien supérieure à celle notée pour le Po-u, puisque, à la fin de la période soumise à examen, 15% de la voix dépasse encore 500 Hz, donc la dynamique reste largement. Nous sommes ainsi en desaccord avec Delack (1978) et Diesteilmann (1982) qui estiment que le Po est stable entre six et douze mois. On est en outre frappé par les variations très brusques du Po qui peut passer plusieurs fois de suite et très rapidement de la zone grave à la zone aiguë inversement. Tout au long de l'année, elle montre des caractéristiques semblables à celles de leur Po-m, à savoir instabilité (E.T. toujours supérieur à 6,5dB). Ils sont tous nettement plus intensives que les vocodes ayant servi à déterminer le Po-u. 1-m. 30dB. Hauteur et intensité évoluent donc de pair.

Po-u et Po-m, avec leurs dynamiques respectives, ne suffisent pas à décrire l'ensemble des possibilités vocales de l'enfant, c'est-à-dire sa tessiture. Le bébé
utilise des comportements laryngés différents de la voix modale : il s'agit essentiellement de deux sortes de productions vocales, le "creak" et le "couinement suraigu" ou "squealing" (respectivement au moins 9 et 4% de l'ensemble des vocalisations). Les creaks sont tous longs, intenses, graves et caractérisés par une instabilité extrême du mouvement glottique qui rend la détection du Fo délicate. Les énoncés suraigu前者 généralement moins nombreux que les creaks peuvent monter jusque vers 1.800-2.000 Hz. Leur intensité est toujours faible, et leur durée brève.

Les deux modes phonatiques ont deux registres de voix qui, selon Hollien (1972), mettent en oeuvre des comportements laryngés complexes. Ceci dénote chez l'enfant des possibilités vocales au moins aussi variées que chez l'adulte.

3. LA VOIX DANS LES INTERACTIONS.

L'analyse permet de dégager une nette différence dans cette situation : en effet, l'enfant n'utilise plus les extrêmes de la tessiture, ni les divers comportements exploratoires. Creaks, énoncés sur aiguères, brusques variations de hauteur disparaissent ; seule la voix modale est employée. Un Fo donné est lié à une modalité linguistique donnée, les extrêmes de la tessiture se situent entre 260 et 760 Hz selon le type d'énoncé ; 85% des émissions se situent en dessous de 500 Hz. On relève aussi, à partir du mois 10 ; un nombre important de chuchotements (10%) absents des deux autres catégories. Il sont tous brefs : une ou deux syllabes. Oller (1980) signalait que ce mode phonatoire n'a pas été relevé dans les travaux sur le langage de l'enfant avant 1 an, mais posait l'hypothèse (1985) qu'il devrait exister. Ce qui confirme nos résultats.

La restriction de la dynamique globale de la voix et l'acquisition du chuchotement témoignent de la rapide socialisation de l'enfant. Celui-ci a appris à "ranger" sa dynamique vocale, et à s'adapter à la situation d'émission. En même temps qu'il produit les premiers énoncés interprétables linguistiquement bien que la couche verbale soit encore absente, il construit sa voix. Celle-ci se structure progressivement, par élimination des registres inadaptés dans la communication sociale, et réorganisation du système initial.

Nos résultats nous amènent à mettre en cause les allégations de la littérature phonologique sur l'évolution du Fo au cours de la première année de vie. Il y est généralement admis que la hauteur de la voix,ainsi que son étendue, augmente régulièrement (Sheppard & al 1968, Papousek 1981) du moins dans les six premiers mois, puis pour les uns, il y aurait stabilité (Diesteimann), pour les autres chute progressive. Or une analyse rigoureuse de la période qui suit nous a permis de remarquer que les [0] déterminant le Fo-u y sont présents. L'enfant aurait-il sa "voix de base" très tôt, les progrès développementaux consistant essentiellement en un élargissement de la tessiture? Mais il est alors surprisant que la maturation physiologique du larynx qui se fait au cours de la première année (accroissement de 100% de la longueur des cordes vocales) n'ait pas plus d'effet sur le Fo-u. Faut-il attribuer ceci au fait que l'enfant domine la régularité des mouvements laryngés dès trois semaines (Fourcin 1978) ? Il agirait donc essentiellement sur la tension des cordes vocales, et compenserait par une augmentation en longueur et en volume. Tout ce point demande encore investigation. Il est cependant clair dès à présent que l'enfant sait, dès le mois, non seulement parfaitement contrôler sa voix, mais encore l'adapter aux diverses situations, et la plier à des contraintes de type linguistique ou social.

**Tab.1**

<table>
<thead>
<tr>
<th>Sujet</th>
<th>A.</th>
<th>B.</th>
<th>C.</th>
<th>D.</th>
<th>E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fo-u</td>
<td>en Hz</td>
<td>en Hz</td>
<td>en Hz</td>
<td>en Hz</td>
<td>en Hz</td>
</tr>
<tr>
<td>8</td>
<td>324</td>
<td>70</td>
<td>220-450</td>
<td>82 %</td>
<td>410</td>
</tr>
<tr>
<td>9</td>
<td>340</td>
<td>74</td>
<td>230-450</td>
<td>85 %</td>
<td>401</td>
</tr>
<tr>
<td>9,5</td>
<td>335</td>
<td>76</td>
<td>260-460</td>
<td>65 %</td>
<td>368</td>
</tr>
<tr>
<td>9,4</td>
<td>335</td>
<td>76</td>
<td>270-460</td>
<td>75 %</td>
<td>402</td>
</tr>
<tr>
<td>10,1+2</td>
<td>324</td>
<td>78</td>
<td>270-460</td>
<td>95 %</td>
<td>359</td>
</tr>
<tr>
<td>10,3+4</td>
<td>354</td>
<td>40</td>
<td>240-480</td>
<td>80 %</td>
<td>341</td>
</tr>
</tbody>
</table>

**Tab.2**

<table>
<thead>
<tr>
<th>Sujet</th>
<th>Fo-u et tessiture Fo-m du Jais</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,1+2</td>
<td></td>
</tr>
</tbody>
</table>

**Fig.1**

**Fig.2**

**Fig.3**

**REFERENCES**

Sheppard W. & al. (1968) in J.Sp. Hearing Res.11, 1, 94.
COMPUTER SYNTHESIS OF INTONATION

K. J. Kohler

Institut für Phonetik, Universität Kiel, Olshausenstr. 46, 2300 Kiel, F.R.G.

ABSTRACT

This paper distinguishes between a global utterance intonation, which is related to the meaning to be conveyed, and local F₀ perturbations, which are due to articulatory constraints. Results are presented of computer synthesis and listening experiments, and a set of rules is given for adjusting the same global contour to different segmental strings in German.

INTRODUCTION

Appropriate F₀ patterns are essential in the synthesis-by-rule of natural-sounding speech. The analysis of natural speech production has established the following points of F₀ control:

1. A global utterance intonation, which is related to the meaning to be conveyed, has to be distinguished from local F₀ perturbations, which are due to articulatory constraints.

2. Apart from local adjustments, F₀ develops as if there were no voiceless sections, i.e., it continues after a voiceless interruption from a value it would have reached had there been voicing.

3. In the presence of voiceless sections, the timing of F₀ has to such that its characteristics (e.g., the peak value and the indication of an F₀ descent) fall within a voiced stretch of speech and can thus be recovered by a listener.

This paper discusses the incorporation of these points into an F₀ synthesis-by-rule for German. It limits itself to falling terminal contours containing a single peak.

GLOBAL F₀

In a sentence such as "Sie hat ja gelen gen." ("She's been lying.") with focus stress on the syllable lo /lo:/, the F₀ peak can be on the syllable ge, preceding the stress, or at the center of the stressed syllable, or at its end (see fig. 1). This shift in the F₀ peak position is correlated with a change in meaning from "established" to "new" to "emphatic".

Using the Kiel Phonetics Institute Speech Signal Processing program (SSP, cf. Schäfer, 1982) and the pitch algorithm in it (Schäfer-Vincent, 1983), LPC-based synthesis of the above sentence with a stepwise shift of the F₀ peak contour A1A2 along the time axis was carried out, starting from the center peak position (see fig. 1a) and moving it to the left in 6 equal steps as well as to the right in 4 equal steps of 30 ms each. The shifted contour was masked in voiceless stretches, its right-hand branch time-expanded in the left shift, and the F₀ transitions smoothed (see fig. 1b). The resulting 11 stimuli were presented once as a series of left to right F₀ peak shifts to 33 native German speakers, who had to mark the positions in the series at which they perceived changes. Table I lists the results.

Table I. Number of changes perceived at the positions II–XI of a series of left to right F₀ peak shifts in "Sie hat ja gelen gen.":

<table>
<thead>
<tr>
<th>Position</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>III</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VI</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VII</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>VIII</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>IX</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>X</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>XI</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

The sharp increase in the first change score from stimulus IV to stimulus V suggests that a category boundary is transgressed at this point in the series. The greater spread of "further change" answers and the lower second maximum score at stimulus X point to a more gradual change after the first category boundary. Stimulus V is the first in the series that has the F₀ peak in the stressed vowel; in stimulus IV it is at the consonant/vowel boundary. In the original stimulus VII, the F₀ peak is located about 100 ms after vowel onset. Since there still is a substantial "first change" score at stimulus VI the F₀ peak should occur later than the 60 ms mark to convey the meaning "new" unambiguously. But it may be shifted into the vowel by another 30 ms without losing its semantic identity because the next maximum score only occurs at stimulus X. Thus the F₀ peak signaling the concept "new" should be located at the center of a phonologically long vowel, i.e., about 100–120 ms into a vowel of 200–250 ms duration at a medium speech rate.

LOCAL F₀

The next question to be answered concerns the manifestation in other segmental strings of an F₀ peak contour with the same meaning "new". To investigate these local F₀ adjustments of the same global contour a set of naturally produced utterances of the type "Sie malt." /*zi 'ma:lt'/, "Sie macht." /*zi 'ma:kt'/, "Sie macht." /*zi 'ma:mət', "Sie machen." /*zi 'ma:xəmən', "Sie strickt." /*zi 'strikt/', "Sie niesen." /*zi 'ni:zən/ etc., i.e., with long and short, low and high vowels in voiced and voiceless consonantal environments and in monosyllabic and multisyllabic words are analyzed with the help of SSP. Then the center F₀ peak contour of "Sie malt." is transferred to the other utterances and adjusted according to the various conditioning factors by the following set of ordered rules.

(1) The F₀ peak is positioned on the time axis in such a way that, within the same speech rate, the timing of the F₀ peak contour as such stays the same (independently of the segmental timing), provided the listener receives a clear indication of the central F₀ rise-fall and its peak value on the stressed word. This means that the F₀ peak occurs at the same absolute time of about 100 ms after stressed vowel onset in all phono-
logically long vowels, and in phonologically short vowels if they are followed by a voiced consonant or another syllable to signal the F0 descent; if, in the latter case, the F0 peak falls inside a voiceless consonant and is thus not recoverable by a listener, it is brought forward to the end of the short vowel. In the case of only a voiceless consonant following a short vowel in a monosyllabic word, 30-50 ms have to be provided for both an F0 rise and an F0 fall inside the vowel, otherwise the terminal nature of the global intonation would not be signaled to a listener.

The whole F0 peak contour in its original duration from the beginning of /a/ to the end of /i/ in "Sie malt" is transferred to the new segmental string and time-locked to the peak point as fixed in (1).

The right-hand branch of the F0 peak contour is time-expanded to fit multisyllabic words.

To account for vowel-intrinsic F0 differences, the F0 peak contour is expanded in the frequency domain for high vowels in such a way that the F0 value at the vowel center (STRESSPOINT) is raised by a constant factor for male and female speakers. The program then interpolates F0 between this point and the time markers at the beginning and the end of the peak contour according to the proportion \( n_1/n \) \( \Delta F0 \), where \( n \) = total number of F0 values between one of the time markers and the stresspoint, \( n_1 \) = number of F0 values from the time marker to the F0 value to be changed. Thus the F0 values at the boundaries of the peak contour stay the same, whereas the others between a boundary and the stresspoint are changed in steps to a maximum \( \Delta F0 \) at the vowel center. This procedure guarantees a maximal intrinsic vowel influence on F0 at the vowel target and takes the preceding and following coarticulation into account, providing weaker effects at vowel onset and offset.

There is an increase of F0 at the transition from a voiceless fricative to a voiced sonorant through a strengthening of the Bernoulli effect, and the opposite at the reversed transition. This consonantal effect on F0 is accounted for by frequency expansion or compression according to the same formula as in (4) with the beginning/end of a vowel after/before a fricative as the stresspoint and the vowel center as a time marker. This way the effect is strongest at the vowel boundary and disappears at the vowel center.

The F0 contour for "Sie" is transferred separately.

F0 is masked in voiceless stretches.

After LPC-based synthesis of the rule-generated F0 patterns over the various segmental strings, the utterances are arranged in pairs, and listeners have to judge whether they represent the same intonation pattern. The procedure of F0 manipulation and auditory evaluation is repeated until listeners accept all the pairs as belonging to the same pitch category. The result is a speaker-independent set of rules which generates all the local F0 perturbations in the same global pattern, starting from one basic contour.

REFERENCES


Fig. 1

(a) Speech wave and fundamental frequency (center peak) in the naturally produced utterance "Sie hat ja gelogen." The end contour (on the syllable gen) was added by F0 parameter manipulation because the analysis did not provide it. The time markers A1, A2 delimit the F0 peak contour (coinciding approximately with /oi/) which was shifted left and right.

(b) The left- and right-most positions of the shifted F0 peak contour on the same time scale as in (a), approximating the natural productions of early and late peaks, respectively.
THE INFLUENCE OF SYNTACTIC STRUCTURE ON FO PATTERNS OF CANADIAN FRENCH SENTENCES

Stephen J. Eady

Centre for Speech Technology Research, University of Victoria, P.O. Box 1700, Victoria, B.C., V8W 2Y2, Canada

The purpose of this study was to investigate fundamental frequency (FO) patterns in declarative sentences of Canadian French containing complex syntactic structures. The results reported here form part of a larger study (Eady, 1983), which also examined French sentences with simple syntactic structures. The aspect of the FO pattern that was studied is called the FO topology --- that is, the peak FO values of the content words (nouns and verbs) of a sentence. Previous work has shown that some complex syntactic structures have an effect on FO patterns in American English (Cooper and Sorensen, 1981). This study was designed to examine this issue for sentences of Canadian French.

METHODS

The procedures used in this study involved the recording of isolated sentences produced by French speakers. The recorded sentences were analyzed using digital techniques and a peak value of fundamental frequency was determined for pre-selected key words in each sentence. Statistical tests were applied to determine the effects of various factors on these peak FO values.

The speakers for this study were 12 undergraduate students (6 males and 6 females) at the University of Ottawa. They were all monolingual speakers of Quebec French and ranged in age from 18 to 22 years. Each speaker was asked to read a list of sentences which were presented in one of six pseudorandom orders.

The stimulus list contained sentences with simple and complex syntactic structures. There were eight simple sentences, each containing four key words whose peak FO was measured. In addition, there were a number of sentences with complex syntactic structures. Among these structures were complement clauses, parenthetical phrases and coordinate clauses. The stimulus list contained four examples of each of these sentence types. Each complex sentence contained key words that were identical to those from one of the simple sentences. The following are examples of each of the four sentence types discussed here (key words are underlined; major syntactic boundaries are denoted with a double slash):

Simple Sentences:
1. La fille de Cécile veut faire du ski.
   (Cécile's daughter wants to go skiing.)
2. Les fils de Pierre vont au cirque près du parc.
   (Peter's sons are going to the circus near the park.)
Complement Clause:
3. Je ne sais si/que la nièce de Cécile veut faire du ski.
   (My sister said that Cécile's niece wants to go skiing.)
Parenthetical Phrase:
   (Peter's sons, Jack cried, are going to the circus near the park.)

Coordinate Clause:
5. La fille de Louise va coudre sa jupe./mais la nièce de Cécile veut faire du ski.
   (Louise's daughter is going to sew her skirt, but Cécile's niece wants to go skiing.)

As indicated above, the stimulus list that was presented to the speakers consisted of these five sentences plus three similar examples for each sentence listed here. The findings reported below are for all examples of each type.

RESULTS AND DISCUSSION

The results for the sentences used in this study are presented in Table 1. This table shows the mean peak FO value for each key-word position of the four sentence types examined here. These values have been averaged across all examples of each sentence type and across all 12 speakers used in this study. A double slash has been inserted in the table to indicate the location of the major syntactic boundaries in each sentence type.

| Mean Peak FO (Hz) for the Key Words of Simple and Complex French Sentences |
|-----------------------------|-----------------------------|
| Sentence Type | Key-Word Position |
|                | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Simple         | 251 | 225 | 219 | 173 |
| Complement     | 249 | 230/227 | 214 | 214 | 172 |
| Parenthetical  | 254 | 256/196 | 235/221 | 171 |
| Coordinate     | 256 | 228 | 221 | 229/231 | 212 | 211 | 171 |

The first thing to notice about the FO patterns of these four sentence types is that they all display declination. That is, in all cases, the peak FO value for the first key word in a sentence is higher than that for the last key word. In fact, all four sentence types have an FO peak of about 250 Hz for the first key word and 170 Hz for the last key word. This results in an overall topline declination of some 80 Hz, regardless of the length or syntactic complexity of the sentence.

However, while the FO patterns for the four sentence types are similar in this respect, they do differ in that the declination pattern is monotonic in two cases, but non-monotonic for the other two types. Thus, for the simple and complement-clause sentences the average FO value at each key-word position is lower than (or equal to) the one before it. For sentences with parenthetical phrases and coordinate clauses, on the other hand, this is not always the case.

The topline contour for the simple French sentences examined here differs somewhat from the pattern observed for comparable sentences of American English (Cooper and Sorensen, 1981). As was indicated above, the simple French topline encompasses a declination of some 80 Hz from the first to the last key word in the sentence. This decline is not evenly distributed throughout the sentence, however, as there is a greater FO drop at the beginning and end of the utterance than there is.
in the middle. The FO decrease between the second and third key words accounts for only 6% (6 Hz) of the total declination observed in these sentences. This archetypical French topline pattern is different from that observed for American English (see Eady, 1984, for further discussion).

The declination pattern observed for simple French sentences is also evident in sentences with complement clauses. Once again, this sentence type has a monotonic FO declination, most of which occurs at the beginning and end of the sentence. The FO drop between the first two and last two key words accounts for almost 80% of the total decline for this sentence type. The FO drop between the intermediate key words is much less, indicating that the topline pattern for complement-clause sentences is similar to the contour observed in these sentences. Most importantly, there does not seem to be any effect on the FO pattern produced by the major syntactic boundary that is present in this sentence type.

For French sentences with parenthetical phrases, on the other hand, the FO topline pattern does differ significantly from that observed in simple sentences. As is evident in Table 1, the FO decline observed for this sentence type is non-monotonic due to a considerable drop in FO between the two key words within the parenthetical phrase (i.e., the third and fourth key words in the sentence). This FO rise amounts to almost 30 Hz, and it is preceded by a drop in FO of even greater magnitude (i.e., an FO drop of 40 Hz between the second and third key words). This pattern produces a relatively low FO value at the beginning of the parenthetical phrase, followed by a sharp rise in FO at the end of the phrase. The resulting FO topline is different from that observed in simple sentences, and this (along with other acoustic cues) serves to demarcate the parenthetical within the main clause of these French sentences. A similar method for demarcating parenthetical phrases has also been observed in sentences of American English (Kutik, Cooper, and Boyce, 1983). However, the shape of the FO topline for English parentheticals is generally falling, in contrast to the rising FO pattern that is observed here for parenthetical phrases in French.

The last sentence type examined here contains two coordinate main clauses and, like the sentences with parenthetical phrases, also exhibits a non-monotonic declination pattern. In the case of the coordinate-clause sentences, however, the departure from the archetypical declination contour is more subtle than that observed for parentheticals. As can be seen in Table 1, the FO topline for this sentence type is declining at all points except between the third and fourth key words. There is a rise of 8 Hz immediately preceding the boundary between the two main clauses (i.e., between key words 3 and 4), followed by an additional increase of 2 Hz after this major syntactic boundary. These results suggest that the coordinate-clause boundary is preceded by an FO continuation rise and followed by a resetting of the FO topline, as has been observed for other major syntactic boundaries in French sentences (Larreur and Emerard, 1977; O'Shaughnessy, 1981). A more thorough examination of the data reveals that all of the speakers used in this study invariably produced a rise in FO either before or after the coordinate-clause boundary and that in some cases an FO rise is evident in both places. Once again, this deviation from the simple declination pattern can be attributed to the effect of the major syntactic boundary in these sentences.

CONCLUSION

Sentences of Canadian French containing three different complex syntactic structures have been compared to syntactically simple sentences with respect to the toplines of their fundamental frequency contour. One of these complex structures (the complement clause) has been found to have a topline pattern similar to the FO contour observed in simple French sentences. On the other hand, French sentences containing parenthetical phrases and coordinate clauses have FO contours that differ significantly from the patterns observed in simple sentences.

These results can probably be attributed to a difference in the strength of the major syntactic boundaries that are present in each sentence type. Based on an analysis using a metric of syntactic boundary strengths (similar to that developed by Cooper and Faccio-Cooper, 1980), it has been found that for the French sentences used in this study the syntactic boundaries associated with parenthetical phrases and coordinate clauses are considerably stronger than the boundary preceding a complement clause. This difference in syntactic boundary strengths is manifested in speech production by a clear distinction between the FO patterns observed for the more complex syntactic structures and those produced for syntactically simple sentences of Canadian French.

ACKNOWLEDGEMENTS

This work was supported by a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada and by a BRS Grant to Hawkins Laboratories. The author is grateful to Arthur S. Abramson, Janet D. Fodor and Patrick W. Nye for their advice and criticisms.

REFERENCES


ACOUSTICAL ASPECTS OF SINGLE VS. DUAL FOCUS IN SENTENCE PRODUCTION

William E. Cooper and Stephen J. Eady

Department of Psychology, University of Iowa, Iowa City, Iowa 52242

Linguistic focus can be defined as the tendency to accentuate or highlight portions of a sentence for reasons related to meaning (e.g., Halliday, 1967). In previous research, we have investigated the acoustical aspects of American English sentences produced with narrow focus (contrastive stress) on one of the lexical items (Cooper, Eady and Mueller, 1985; Eady and Cooper, 1985). In this paper, we report recent work used to elicit acoustical analysis for sentences containing two focused words. Our aim here is to determine whether the addition of a second focus produces any contextual effects on intonation that are not attributable to localized single-focus influences.

Sentence Materials

The base stimuli for this study consisted of six sentences of English with simple syntactic structures, each containing three key words for which F0 and duration were measured. All key words were nouns and were chosen so that they could easily be segmented from adjacent words on a digitized waveform display. An example, with key words underlined is "Don shot the puck to Kent."

For this study, a set of four priming questions was composed for each base sentence. For example, the questions used with the example sentence above were as follows:
A. What happened?
B. Who shot the puck to Kent?
C. Who did Don shoot the puck to?
D. Who shot the puck to whom?
These stimuli were used to elicit four versions of each base sentence. The versions corresponded to non-focused, sentence-initial focused, sentence-final focused, and dual-focused conditions, respectively.

Subjects, Recording and Analysis Procedures

The subjects for this study were seven male undergraduate students at the University of Iowa. All were native speakers of American English with no apparent hearing or hearing impairments. Six speakers were tested individually in a quiet room. Each speaker was seated in front of a microphone and given a stack of file cards with a sentence typed on each card. On each trial, the speaker listened to a prerecorded priming stimulus presented via headphones and then read the sentence aloud into the microphone with the appropriate emphasis pattern. The speaker's responses were recorded on another real-to-real tape recorder at a speed of 3.25 inches per second.

At the start of the recording session, each subject was presented with six practice sentences to familiarize him with the recording procedure. The 24 target sentences (6 base sentences x 4 versions) were then presented in one of two pseudorandomized, counterbalanced orders, along with 58 filler sentences which were used to separate the different versions of the target sentences.

The recorded sentences of each speaker were subjected to a perceptual evaluation, to ensure that they had been produced as intended. A phonetician blindly assessed all productions of the seven speakers and judged the sentences to have either sentence-initial, sentence-final, dual or neutral focus. The perceptual evaluation revealed a total of only six production errors for the seven speakers. The average number of correctly identified target sentences per speaker was 23.1 out of a total of 24.

The acoustical analysis for the recorded sentences used digital techniques to measure peak F0 for all key words, and duration for the first and last words in each sentence. We then tested for significant differences in the acoustical patterns, using standard statistical procedures (see Cooper et al., 1985, for details).

Results and Discussion

The results of the acoustical analyses for the sentences of this experiment are presented in Table 1, where the means are arranged according to sentence version and key-word position. Each value in the table represents the mean for all six sentences averaged across all seven speakers. These results are presented graphically for duration in Figure 1 and for fundamental frequency in Figure 2.

Table 1

<table>
<thead>
<tr>
<th>Mean F0 and Duration of Key Words Averaged across Six Sentences Produced by Seven Speakers (Values in parentheses are standard deviations)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key-Word Position</strong></td>
</tr>
<tr>
<td><strong>Version</strong></td>
</tr>
<tr>
<td><strong>Neural F0</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
<tr>
<td><strong>Initial F0</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
<tr>
<td><strong>Final F0</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
<tr>
<td><strong>Final F0</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
</tr>
<tr>
<td>(msec)</td>
</tr>
</tbody>
</table>

Duration

As we have found in previous studies, the results of this experiment indicate that focusing a word yields an increase in duration, and that the magnitude of this increase is greater in sentence-initial than in sentence-final position. Furthermore, we observe no difference in the extent of durational increase between single-focus and dual-focus sentences. These trends are evident in Figure 1.

At the first key word position, we find a significant difference among the four sentence types (F(3,33)=19.02, p<0.005), due to the increased durations for the initial-focus and dual-focus versions. The amount of durational increase for these two versions is equivalent to 34.4% (initial focus) and 46.3% (dual focus) of the mean duration for the same words in the neutral version. This magnitude of durational increase is very similar to the 35% to 45% range that was observed for sentence-initial focused words in a previous study (Cooper et al., 1985).

The durational pattern for sentence-final words in the present experiment is also similar to our previous results. Once again, we find that focused words in this position (i.e., the final-focus and dual-focus versions) have significantly longer durations than unfocused words (i.e., the initial-focus and neutral versions; F(3,31)=23.62, p<0.005). In this case, however, the amount of
dursional increase due to focus is smaller than in sentence-initial position, ranging between 13.3% (final focus) and 15.3% (dual focus) of the mean duration for the neutral version. This smaller amount of elongation for sentence-final words is very similar to the results of our previous work, and can probably be attributed to an interaction with the segmental lengthening that is typically observed in this position (see Cooper et al., 1985, for further discussion of this point).

Fundamental Frequency Topline

The $F_0$ patterns presented in Figure 2 show the influence of sentence focus on the production of these utterances. Analyses of variance calculated at each key-word location for the sentence focus. We have found no significant difference in peak $F_0$ at all three sentence positions.

At the first key-word position, the significant $F$ difference among the four sentence versions ($F(3,28)=1.88, p<0.05$) is due to the fact that the initial-focus and dual-focus versions average almost 14 Hz higher than the two versions with no focus on the first key word. This bimodal distribution among the four sentence types holds for six of the seven speakers and five of the six sentences examined here. This finding stands in contrast to the results of our previous studies of focus with longer sentences (i.e., Cooper et al., 1985; Eady and Cooper, 1985), in which we found no significant effect of this kind on the $F_0$ peak of the first key word.

The peak $F_0$ values for the second key-word location in these sentences are also affected by the position of a second focus word. The peak $F_0$ difference in $F_0$ among the four versions is significant ($F(3,23)=10.75, p<0.005$). In this case, however, the difference is due to the low $F_0$ value for the initial-focus version, which is significantly lower than that of the other three sentence types. This pattern of results holds for all speakers and sentences examined here, and coincides with our findings from previous studies. The new finding from the present experiment is that the $F_0$ value for the second key word in the dual-focus version does not drop as low as that of the initial-focus version. Thus, even though both sentence types have focus on the first key word, the dual-focus sentences do not have the low flattening typical of the other sentence types. This pattern demonstrates that the sentence-final focus is similar to that of the neutral and sentence-initial focus versions.

The lack of a lowering of $F_0$ peak for the second key word in the dual-focus version seems attributable to the influence of the sentence-final focused word in this sentence type. As can be seen in Figure 2, the last word in the dual-focus version has a $F_0$ peak that is higher than that of the sentence with initial focus. The $F_0$ values for these two sentence types are significantly higher than the average frequencies for the third key word in the neutral and initial-focus versions ($F(3,31)=21.56, p<0.005$). This increase in $F_0$ at the end of the dual-focus version would be facilitated by a minimizing of a frequency drop in the middle of the sentence compared to the $F_0$ pattern of the initial-focus sentences.

Conclusion

In summary, the results of this experiment indicate how the acoustical patterns of English sentences are influenced by sentence focus. We have observed that the placement of narrow focus on a lexical item results in a heightening of the $F_0$ peak and an increase in duration for the focused word only. Our new findings indicate that this pattern is evident regardless of whether the sentence contains one or two focused items. While there is no direct acoustical interaction between two foci in the same sentence, we have observed an effect on the $F_0$ peaks of words between the two focused items. Such words do not have the lowered $F_0$ values that are observed when they occur in utterances with a single focus in sentence-initial position. We have attributed the lack of post-focus $F_0$ drop in dual-focus sentences to the presence of the additional focused item at the end of the utterance. With the exception of this anticipatory effect on unfocused words, however, we see no interaction between two narrow-focused items in the same sentence. Thus, the acoustical attributes of narrow focus are relatively constant, regardless of whether a sentence contains one or two foci.

References

A FAST SPECTRAL COMB ALGORITHM FOR Fo DETECTION

Ph. Martin
Institut de Phonétique, Université de Provence
France

Since the recent availability of fast and moderately priced hardware, fundamental frequency detection based on spectral analysis is becoming more and more popular. Using more extensively signal related information than time-domain approaches to detect Fo, these methods appear to be resistant to noise and, to a larger extent, to the non-stationary nature of the speech signal.

**The spectral comb method**

The spectral comb method was introduced recently as another method based on Fourier analysis for reliable Fo tracking and voicing detection. To evaluate Fo, the short time power spectrum of an appropriate windowed signal frame \( F(w) \) is first "trimmed" by replacing spectral peaks by narrow parabolas when certain selection criteria are met, and by zeroing the remaining of the spectrum. This ensures that non-harmonic related values (in particular spectral valleys) will not interfere in the overall computation.

![Graph](image)

"Trimmed" Spectrum: all peaks above an amplitude threshold are replaced by truncated parabola.

The trimmed spectrum \( F(w) \) is then cross-correlated with a spectral comb function \( C(wp, w) \) with teeth of decreasing amplitude and variable intervals \( wp \).

\[
C(wp, w) = \sum A_n \delta (nwp - w)
\]

with \( A_n = \) amplitude of the \( n \)th tooth. Almost any decreasing function of \( n \) can be used to define the teeth amplitudes. Using for instance \( n \exp (1/8) \), the cross-correlation \( I(wp) \) becomes:

\[
I(wp) = \sum \exp (1/8) | F(wp) |
\]

The maximum of the cross-correlation function \( I(wp) \) is reached when a large number of the comb's teeth coincide with the harmonic peaks of the spectrum. When this value exceeds a voicing threshold, the corresponding tooth interval is taken as Fo.

The use of truncated parabola ensures that rounding errors due to truncation of the frequency values will not lead to missing harmonics in the cross-correlation.

![Graph](image)

Spectral Comb Function with teeth of decreasing amplitude.

**A faster method**

The computational effort involved in the implementation is roughly comparable to the cepstrum.

If \( F(w) \) is represented by \( m \) values, and \( n \) the number of comb's teeth, \( n \cdot (m^4) \) operations are necessary to compute \( I(wp) \). For example, a 1 kHz bandwidth Fo search, with a 4 Hz resolution, would require 15,625,000 products and accumulations once the Fourier transform is computed.

The figures above show clearly that many of the spectral and comb values involved in the cross-correlation are equal to zero. A much more efficient algorithms can therefore be obtained by involving only non-zero spectral and comb values.

Instead of computing \( I(wp,w) \) as a sum of products, it would be easier to include only the non-zero factors of both spectral functions. This can be obtained by considering the cross-correlation as the sum of harmonic parabola scaled on the frequency axis by the rank of each comb tooth, and multiplied by the amplitude of the tooth.
With a logarithmic frequency scale, the computation gain is independent from the number of spectral peaks. For a 1000 Hz range and a 4 Hz resolution, 2 · 250 additions are necessary for each spectral shift, leading to a total of 5 · 2500 = 5000 sums (or subtractions) for 5 spectral shifts.

References


The comb cross-correlation function is the sum of frequency scaled trimmed spectrum, with an amplitude reduced by the comb tooth order.

The exact computational gain will depend on the number of peaks in the spectrum, but can be typically estimated as 6. If n values are used to represent each spectral parabola, and p peaks were detected, n · p sums and products are necessary to evaluate the cross-correlation function I(wp).

Thus if p = 10 (10 spectral peaks), the overall computation effort is reduced to 40 divisions for frequency shifting and 1,650 additions if each parabola is sampled by 33 values.

A logarithmic spectral comb method

If a logarithmic scale is used for both the amplitude and the frequency axis, a further improvement of the algorithm can be obtained. Instead of being scaled down on the frequency axis according to the order of the comb's tooth considered, the corresponding trimmed spectra are merely shifted along the frequency axis by factors of log n (n being again the order of the tooth) and added.

Thus, if a logarithmic scale is used, the cross-correlation function is obtained by iteratively summing, for each spectral peak, a series of parabola:

- shifted on the frequency scale by factors of 1, 1/2, 1/3, ..., 1/n according to the comb tooth order n;

- shifted on the amplitude scale by a factor of 1 dB, 2 dB, ..., n dB according to the comb tooth order n.

Each shift reduces the trimmed spectrum in frequency and in amplitude according to the tooth order. The sum of those reduced spectra corresponds to the cross-correlation function whose maximum defines the fundamental frequency.
COMPARING DIRECT SPECTRAL MATCHING TECHNIQUES WITH FORMANT EXTRACTION UNITS FOR SPEAKER RECOGNITION

A. Paoloni, A. Federico*, C. Iba

Fondazione "Ugo Bordoni" Viale Trastevere, 108
00153 Roma, Italy

* ENEA, Casaccia - Via Anguillarese 1300
Roma, Italy

The implementation of a reliable system for the text independent speaker recognition on telephonic channels is of great concern in several fields ranging from military to banking operations and, of course, for forensic applications [1].

The Authors developed a system based on the extraction of an extended set of acoustic parameters from the Italian vowels. They were essentially the formants and the fundamental frequencies of each vocalic sound. A statistical model and a pattern classifier were derived and verified on a very large database of realistic acoustic signals [2,3].

The experience with this recognition system suggested a very favourable performance rate with this method, even over disturbed telephonic channels. The condition was a great care to be devoted to the vocal segments extraction and to the spectral measurements. This way the operator intervention and its technical skill became the crucial feature of the recognition system and a serious trouble for every real time application.

The first bottleneck to be overcome is the manual extraction of the formant frequencies that actually is a computer assisted operator job. A subsystem, named ARPS (Analysis, Representation and Elaboration of Speech), works on a VAX 11 computer with an high level colour CRT (Tektronix 4115), and allows the parameters interactive measurement under operator control.

The complete procedure automation would require the automatic segmentation of the signal, followed by the automatic labeling of acoustic frames by means of a statistical clustering algorithm. In this work we present only the results of the substitution of the formant shape parameters by the formant frequencies extracted by manual procedure. It is well known that the Linear Prediction Coefficients and the first Cepstral lines are good estimators of the resonance spectra of the speaker voice.

EXPERIMENTAL SETUP

To have a realistic comparison among the proposed parameter sets a five speaker telephonic database was collected reproducing an experimental environment close to the real life.

A number of 820 acoustic frames was selected from vowels in the speech by listening and signals displaying on the ARPS. For each frame 8 LPC coefficients, 16 cepstral lines and 3 formants along with the pitch measure were either computed or directly measured. A complete parametric set was composed by association of the four Italian vowels parameters so giving features vector of dimensionality 16 for the formants, 36 for the LPC and 68 for the cepstrum associated to any speech release from the 5 speakers. Final vectors were derived by average on five of those measured vectors and then used in a statistical classification experiment.

The ultimate quality factor of such an experiment is the percentage of false identifications (FID) and false rejections (FNR). Unfortunately the first kind of error is strongly dependent from the choice of the speakers, if the speakers number is limited. Therefore to improve the significance of the merit factors we compute also the ratio (δ) of the average between speakers distance. To avoid the limitation of such a purely geometrical scattering index a statistical distance definition was adopted [1].

In fig. 1, 2 and 3 non linear maps of the three complete distance matrices show an approximate view of the three parameter sets properties. In each figure, the 5 tested speakers are represented by 5 different symbols.

![Fig. 1 - Representation of distance matrix computed by 16 formant parameter sets](image)

![Fig. 2 - Representation of distance matrix computed by 36 LPC coefficients](image)
In Tab I the scores of $\varepsilon_1$, $\varepsilon_2$ and $\delta$ are quoted. The main result of the experiment is that the perceivable worsening of the manual to automatic feature extractor does not significantly affect the classification quality of LPC and Cepstrum in terms of error rates. Therefore the adoption of an automatic feature extractor is a viable choice toward a completely automatic but reliable speaker recognition system.

<table>
<thead>
<tr>
<th>INPUT STATISTICS</th>
<th>variates 16 formants</th>
<th>variates 32 LPC</th>
<th>variates 64 cepstrum</th>
</tr>
</thead>
<tbody>
<tr>
<td>test conf. level</td>
<td>$\varepsilon_1$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>test conf. level</td>
<td>$\varepsilon_2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>test conf. level</td>
<td>$\varepsilon_1$</td>
<td>0</td>
<td>5.5</td>
</tr>
<tr>
<td>test conf. level</td>
<td>$\varepsilon_2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MERIT FACTOR</td>
<td>$\delta$</td>
<td>12.1</td>
<td>4.3</td>
</tr>
</tbody>
</table>

Tab. 1 - Performance evaluation of speech parameter sets
ONSET TIME AND THE DETECTABILITY
OF LONG DURATION TONES

S. K. McFadden

Defence and Civil Institute of Environmental Medicine, 1133 Sheppard Ave. W., P.O. Box 2000, Downsview, Ontario, Canada, M3M 3B9

INTRODUCTION

Most studies on signal detection use short duration sounds (less than a half second) with rapid onsets. Experiments using long duration sounds can be tedious to run and participate in. Observers find it difficult to pay attention for the durations involved. Moreover, since speech is characterized by short duration rapidly changing sounds, the detectability and discriminability of similar types of sounds has often been more interesting. However, certain tasks, such as those carried out by sonar operators, require the detection and discrimination of long-duration, slowly changing sounds. The few studies that have been carried out on duration indicate that detectability does not improve very much once duration exceeds a second (Green, Birdsell, & Tanner, 1957). Even these studies typically use sounds with very short onsets. Thus the results may not be relevant to the sonar world. Experience suggests that the ear is better at detecting changes in auditory input than in monitoring steady states. Since the ear appears to tune out unchanging sounds, it may be that long duration sounds with slow onsets are less detectable than sounds that change rapidly.

The present set of experiments was aimed at studying the detectability of long duration sounds as a function of onset. In experiments using short duration sounds with short onsets, duration is usually measured from onset to offset. When using long duration sounds with long onsets, it may be more appropriate to consider the overall duration of the sound or the time the sound is at maximum amplitude. In the first experiment, detection was measured as a function of onset time with duration from onset to offset fixed. In the second experiment, detection was measured as a function of onset time with duration at maximum amplitude fixed. In all the conditions, onset and offset times were the same.

METHOD

Subjects

Ten observers, ranging in age from 20 to 50 participated over the course of the study. All had normal hearing.

Apparatus

The overall running of the experiment and signal generation was controlled by a PDP-11/34. Signals were calculated in advance, stored on a disk and played back at runtime at a sampling rate of 8 KHz. The broad-band masker was generated using a B&K model 1405 noise generator, attenuated, and mixed with the signal at the tape recorder. The signal and masker were presented monaurally to the right ear. A VT100 (Digital Equipment Corporation) video terminal was used for presenting non-auditory information and for observers to record their responses.

Stimuli

The signal was a 500 Hz tone. In the first experiment, signals in a given condition had a duration from onset to offset of 3 seconds and a onset and offset time of 0.0005 (signal turned on at zero phase), 0.01, 0.1, 1 or 2 seconds. Thus, overall signal length varied from three to five seconds. In the second experiment, signals in a given condition had a duration at maximum amplitude of 1, 2 or 4 seconds and an onset and offset time of 0.001, 0.01, 0.1, 1 or 2 seconds. Overall signal length ranged from 1 to 8 seconds. In both experiments, signal level during a run was varied according to the rules of Pest (Taylor and Creelman, 1967). Pest is an adaptive technique designed to determine rapidly and accurately the stimulus level, in this case signal intensity, required to achieve a certain performance level, in these experiments 80% correct.

The masker was a continuous broad-band noise fixed in amplitude at 20 dB/Re SPL.

Procedure

Observers were seated in a soundproof booth. Their task was to state in which of two intervals a signal had been presented. A trial consisted of a 300 millisecond warning interval, two signal intervals separated by a half second, and a one second response interval. The length of the signal intervals corresponded to the overall length of the signal being tested during the run. Observers could respond at any point during the trial. Feedback was presented visually as soon as a response was made.

RESULTS

In the first experiment, the signal level resulting in 80% correct performance (Figure 1) increased approximately 3 dB as onset increased from 0.0005 to 2 seconds. As seen in Figure 2, all of the observers showed a similar performance decrement as onset time increased. Based on these results, it would appear that detection is affected by onset with long onsets impairing detectability.

![Figure 1: Average signal level leading to 80% correct as a function of signal onset/offset when signal duration from onset to offset is constant at three seconds. The standard deviations for each condition are shown also.](image-url)
FIGURE 2: Average signal level leading to 80% correct detection as a function of signal onset/offset for each of the five observers participating in the first experiment.

The purpose of the second experiment was to determine the effect of onset when signal duration at maximum amplitude was kept constant rather than signal duration from onset to offset. The results of this study are shown in Figure 3 for the 4 seconds at maximum amplitude condition. Except for possibly one observer (shown by unfilled squares), detection performance was not affected by onset. Moreover, performance in this experiment was more variable both across and within observers. Performance at any one onset differed as much as 7 dB. Initially, it was thought that the increased variability might be due to the wide range of durations used in the second experiment. To check this hypothesis, the four second condition was repeated at the two shortest and two longest onsets. The results were similar to the original results for the four second condition. The means in Figure 3 are averaged over all of the 4 second data.

DISCUSSION

The reasons for the differences between the two studies are not clear. It was not due to keeping the duration at maximum amplitude constant, since the results for the second experiment were examined as a function of onset, with signal duration from onset to offset fixed at four seconds; there was still no effect due to onset.

The simplest explanation is that the differences between the two experiments are due to population differences. Only one observer (shown by open squares in Figures 2 and 3) participated in both experiments. Except for the shortest onset, that observer’s results were similar in the two experiments. Presumably, the two sets of observers were using different cues to detect the signals. The results of the first experiment suggest that those observers were relying at least in part on changes in amplitude. The cues being used by the second set of observers are less clear. However, comparing the results in Figures 2 and 3, the cues used by the second set of observers can result in superior detection performance. In the first study, only 16% of the stopping levels were below 30 dB SPL. In the second experiment, 28% were below 30 dB SPL.

One additional factor should be noted. The current experiments used a 2AF methodology and an adaptive psychophysical technique. Both of these tend to reduce variability and performance decrements due to variation in attention (as compared to fixed-level, yes-no experiments). The observers knew that a signal would occur in one of two specified intervals. Second, if they lost track of the signal, intensity was increased until the signal was reliably detected. Both of these could have reduce the expected effect of long onsets.

CONCLUSION

In general, it would appear that onset can affect the detectability of long duration sounds, but that the effect is observer-dependent. Since the basis for the differences in performance across observers is not understood, further study is required.

REFERENCES


AUDITORY TEMPORAL INTERACTIONS AND FREQUENCY SELECTIVITY

Gregory J. Fleet and John C. Booth

Departments of Psychology and Communicative Disorders, University of Western Ontario, London, Ontario, CANADA, N6A 5C2

Introduction

There has been a long standing interest in the auditory system's frequency selectivity as revealed by tone masking studies. The results are argued here to provide a picture of how the auditory system might process spectral information.

Recently, attempts have been made to examine processing with slightly more complex stimuli or presentations. For example, introducing a second pure tone simultaneously with a pure tone masker reveals a strong "masking" of the signal, due to the interactions of suppression between the two masker components (Shannon, 1976). Likewise, studies where two sequential sinusoidal maskers precede a pure tone signal have shown that more masking results than would be predicted by energy summation of the maskers (Neff, 1984).

A third attempt at modifying this paradigm has been carried out in our lab. The data presented in this paper are taken from a group of studies with a pure tone masker and signal, and where a second low-level pure tone was added to the signal. Specifically, we were interested in predicting the combined masking pattern (i.e., with a double-frequency probe) using the single-frequency patterns. There is evidence, though, that additional factors may play a role in the accuracy of such predictions. In single auditory nerve fibers, the time structure of the response to pure tone pairs reveals strong temporal interactions between each component (Rose et al., 1971). For example, these time patterns show evidence of the auditory system's nonlinear response (two-tone suppression, combination tones), as well as a differential sensitivity to the pair compared to the response of each tone separately (Arthur, 1976; Rose et al., 1974). Therefore, the ability of the single-frequency masking patterns to predict performance may depend on the usefulness of this type of coding.

EXPERIMENT I

Method

A forward masking procedure similar to the psychophysical tuning curve paradigm was utilized. The masker levels were adapted, in a staircase procedure, to just mask the presence of the brief (10 ms) low-level probe. The probe was comprised of one or two pure tones. The masker frequencies were examined ranging from 1.0 to 1.5 kHz. The two probe frequencies were set at 1.0 and 1.5 kHz. The first experiment compared the masking of the probe pair at three amplitude ratios, corresponding to 0, 5, and 10 dB differences in the probe levels.

Results and Discussion

The first condition tested the forward masking of the single-frequency probes at 10 dB SL, and the double-frequency probe components at 10 dB SL each. The results, for one subject, showed an increased detectability of the probe pair compared to masker level necessary to mask each component in the probe, but only by the amount expected by the 3 dB increase in the peak amplitude of the double-frequency probe waveform. This finding, then, would be predicted on the basis of a simple model of single-frequency probe masking patterns. A second subject showed no difference in performance.

When masking levels were obtained for the probe conditions with different amplitude ratios, though, significant differences were found. For example, in Figure 1 the masker levels are plotted as a function of three masker frequencies for a 1.0 kHz probe at 10 dB SL (filled triangles), and a 1.5 kHz probe at 5 dB SL (filled diamonds). One would expect that the masking pattern for the combined probes would approximate the greater of the two single-frequency probe patterns; that is, the filled diamonds at 1.0 and 1.1 kHz and the filled triangle at 1.2 kHz. When both components are presented simultaneously (open squares), an average of 5 dB more masker energy was necessary to mask the pure tone pair. A second subject also showed a significant difference, but only with the 1.1 kHz masker.

Figure 1. Masker levels for single- and double-frequency probes at 10 and 5 dB SL

In the third condition, with a 10 dB difference in the probe levels, even larger masker level differences were found. Figure 2 shows the masker levels for a 1.0 kHz 10 dB SL probe (filled triangles), as well as the approximate levels for the 1.5 kHz 0 dB SL probe (filled diamonds -- Note: These data points show large variability across separate trials; a standard error of approximately 20 dB). Interestingly, when both frequency components are presented simultaneously in the probe duration, masker levels increased substantially, especially for the 1.0 kHz masker.

Figure 2. Masker levels for single- and double-frequency probes at 10 and 0 dB SL
Again, based on the single-frequency probe pattern model, one would expect the results to parallel or coincide with the data shown by the filled triangles. (Two other subjects, though, have shown similar effects; e.g., with the 1.0 kHz masker a 15 to 20 dB masker level difference between the single- and double-frequency probe conditions.) Also, when the levels of the two probe components are reversed (i.e., 1.0 kHz at 0 dB SL and 1.5 kHz at 10 dB SL), significant masker level differences (approximately 10 dB between the single-frequency pattern and the double-frequency pattern) were found for three masker frequencies (1.3, 1.4 and 1.5 kHz).

We can note, then, that a simple model of predicting the double-frequency masking pattern from the single probe data is, in some conditions, not very accurate. Therefore, combining two frequencies in the probe using a pure tone forward masking task, improves the detectability relative to the single masking levels in the single probe conditions. These data are difficult to account for. Our present interpretation is that the improved performance is due to a sensitivity to the temporal coding of each component in the double-frequency probe. For example, physiological studies of single auditory nerve fiber response show evidence that the time structure is a highly sensitive indicator of the sound pressure levels of each tone in the pair (Rose et al., 1974). Therefore, these data may reflect a behavioral manifestation of such coding.

**EXPERIMENT II**

The possibility remains, though, that low-level combination tones may be present in this paradigm, produced by the two simultaneous probe components. (In fact, the cubic difference tone and simple difference tone are equal in frequency with this stimulus configuration at 0.5 kHz.) Although the frequency separation and the absolute levels of the probe components make the presence of combination tones unlikely, a second experiment sought to confirm this idea.

**Method**

Since combination tone generation requires a monaural presentation of the primaries, the first experiment was simply replicated with a dichotic presentation of each probe component. Therefore, the masker was presented binaurally, the 1.0 kHz probe to the subject's right ear, and the 1.5 kHz probe to the left ear. Masker levels were again obtained for the binaural masker with the single monaural probes or dichotic probes.

**Results and Discussion**

Figure 3 shows the data from one subject for the stimulus configuration of a 1.0 kHz 1.0 kHz probe and a 0 dB SL 1.5 kHz probe. As can be seen from the data, the relative difference in performance between the single- and double-frequency probe conditions remains. Further, when the probe levels were reversed (i.e., 1.0 kHz at 0 dB SL and 1.5 kHz at 10 dB SL), the results replicated those found in Experiment I. A second subject also showed similar masker levels with these stimulus conditions. Therefore, even when the possibility of combination tones is removed, greater masker levels are necessary to mask the double-frequency probes relative to the single-frequency conditions, suggesting that performance differences must be due to additional information.

**SUMMARY**

The amount of masker energy necessary to mask single- and double-frequency probes was measured for three probe amplitude ratios. Masker levels greater than predicted by the single-probe patterns were observed with two of the amplitude ratios. This differential detectability of the probe pair is unlikely due to the energy summation of the tones, especially when the masker levels show increases of 15 to 20 dB. Further, these differences remained when a dichotic presentation of the probe was used. This eliminates the possibility of the subjects utilizing energy information at the combination tone region. The data suggest, then, that aspects of the temporal encoding of the waveform may be useful for indicating the amount of energy at each frequency region of the probe. In general, they show that measures of frequency selectivity with more complex stimuli cannot be predicted by measures obtained with simple stimuli.

**References**


FREQUENCY DISCRIMINATION OF TONES PRESENTED IN VARIOUS NOISE BACKGROUNDs

David S. Emmerich, Deborah A. Fantini and William S. Brown

Psychology Department, State University of New York at Stony Brook, Stony Brook, New York 11794, U.S.A.

INTRODUCTION

Bekesy's classic studies demonstrated that even a simple auditory stimulus such as a pure tone of moderate intensity creates a broad pattern of activity on the basilar membrane. It was pointed out that the resulting neural activity in the eighth nerve involves a large population of fibers. If two tones differing slightly in frequency are presented successively, it is also known that the corresponding patterns of eighth nerve action potential sweeps from many name fibers. Such overlap in the populations of active fibers might be expected to make it quite difficult to tell the tone apart; however, human frequency discrimination performance is actually quite good.

In this regard Bekesy (1967) proposed that lateral inhibitory interactions act to sharpen the broad excitation patterns observed in the auditory periphery (Emmerich, Fantini, and Navarro, 1983). This sharpening would make the excitation patterns more discriminable by emphasizing the peaks of patterns of activity at the expense of areas of lesser activity. In contrast, Whitfield (1967) proposed a very different mechanism for frequency discrimination in its model. The information from the tails of patterns of activity is, in fact, especially important in that it signals a shift in the locus of activity with changes in frequency.

With these different possibilities in mind, Emmerich, Brown, Fantini, and Navarro (1983) set out to investigate the use of information from frequency regions remote from the nominal frequencies of the tones, when subjects are given a frequency discrimination task. In this study the tones to be discriminated were detectable and were presented in the centers of the "notches" of a series of band-reject noise backgrounds of different notch widths. As the width of the band of rejected frequencies was varied, it was expected that there would be a corresponding variation in the degree of disruption of the information from the tails of the excitation patterns associated with the tones. For comparison purposes, subjects were also given a signal detection task in which the tones to be detected were presented in the same backgrounds as were used during frequency discrimination. The results indicated that subjects could not detect the detection task information over a wider frequency range when given the discrimination task than when given the detection task. The research described below extends this work to other stimulus situations in an attempt to gain a better understanding of the processing of frequency information.

EXPERIMENT I

By presenting the tones in the center of the band of rejected frequencies, Emmerich et al. (1983) disrupted information from both tails of the corresponding excitation patterns at the same time. In Experiment I the tones were presented at frequencies higher and lower than this band in order to determine if information from the high and low-frequency tails is useful in frequency discrimination.

Procedure

Three experienced listeners served as subjects. On each trial two tones were presented and the subject indicated which was higher in frequency. The frequency separation between the tones was varied across trials using a version of an adaptive forced-choice procedure (Levit, 1971) which estimated the separation which would lead to 75 percent correct responses. This value was taken as the difference limen (DL). Tones were 200 ms in duration and had an intensity of 60 dB SPL. Standard frequencies of 250, 500, 1000, 1500, 1750, and 2000 Hz were employed.

Results and Discussion

To the extent that information from the tails of the excitation patterns is useful in frequency discrimination, the task would be expected to be easier if the tails fell in the notch of the band-reject noise. That is, the DLs obtained in band-reject noise should be significantly smaller than those obtained in the white noise background. This was the case: for the 250 and 500 Hz standards, the mean DLs improved in the band-reject noise by 11 and 4 percent, respectively. For the 1500, 1750, and 2000 Hz standards the improvements were 19, 10, and 33 percent. (By comparison, for the 1000 Hz standard, which was presented in the center of the notch, the improvement was 55 percent.)

Thus performance was facilitated both for tones whose frequencies fell below the notch and for those whose frequencies fell above the notch, indicating that the high- and low-frequency tails each convey information which is useful for frequency discrimination.

EXPERIMENT II

Since discrimination performance in a white noise background varies with frequency, and since different portions of the tails of the excitation patterns of the various tones were released from masking in the band-reject noise of Experiment I, it is difficult to draw any firm conclusions regarding the relative usefulness of information from the low- and high-frequency tails of the excitation patterns from that experiment. In order to address this question, in Experiment II performance was investigated for a single standard frequency presented close to the edge of a high-pass noise, close to the edge of a low-pass noise, and in various combinations of low- and high-pass noise with different frequency cutoffs.

Procedure

The same three experienced listeners served in Experiment II as had served in Experiment I, and the same adaptive forced-choice procedure was used to assess frequency discrimination performance. In the second experiment a fixed high-pass noise with a frequency cutoff of 935 Hz and a fixed low-pass noise with a frequency cutoff of 1056 Hz, plus other low- and high-pass noises with various frequency cutoffs, were used. The standard tone had a frequency of 993 Hz (this is the geometric mean of the cutoffs of the fixed high- and low-pass filters and is approximately 60 Hz into the passband of each of them). The tones again had a duration of 200 ms.
and an intensity of 60 dB SPL.

In the first set of conditions the fixed high-pass noise was presented by itself or together with low-pass noise with a cutoff of either 200, 500, or 800 Hz. In the second set of conditions the fixed low-pass noise was presented by itself or together with high-pass noise with a cutoff of either 2500, 2000, 1500, or 1250 Hz. (In each case there were also control conditions in which the tones to be discriminated were presented in the quiet, and in unfiltered white noise.) Each noise had a spectral level of 35 dB SPL per cycle in the passband.

Results and Discussion

It was anticipated that comparing the DLs in fixed high-pass noise alone with those in fixed low-pass noise alone would assess the relative usefulness of the two tails of the excitation patterns, and that combining additional (low-pass or high-pass) noise with the fixed (high-pass or low-pass) noise background would allow the processing of information in the tails. It was expected that the addition of more noise would make the discrimination task increasingly difficult, and that it would be possible to examine the range over which information is integrated separately for each tail.

The DLs in fixed high-pass noise alone were significantly smaller than those in fixed low-pass noise alone, which is consistent with the hypothesis that the low-frequency tail is more useful for frequency discrimination than is the high-frequency tail. When low-pass noise was added to the fixed high-pass noise, the DLs did become larger, on the average, but they did not increase in an orderly fashion; in high-pass noise alone the mean DL was 1.89 Hz, and adding low-pass noise with cutoffs of 200, 500, and 800 Hz changed the mean DLs to 2.32, 1.83, and 2.43 Hz, respectively. (Performance in unfiltered noise was considerably worse; here the mean DL was 3.94 Hz.)

The results obtained with the fixed low-pass noise were more surprising. Unexpectedly, the mean DL of 4.21 Hz in fixed low-pass noise alone was even larger than the value of 3.94 Hz obtained in unfiltered white noise. Furthermore, adding the high-pass noise with cutoffs of 2500, 2000, 1500, and 1250 Hz did not further impair performance: if anything, the DLs actually decreased: the corresponding means were 4.14, 2.98, 3.74, and 3.15 Hz. Experiment III was undertaken in order to investigate a factor which may have contributed to the unexpected results of Experiment II: the sharpness of spectral edges near the tones to be discriminated.

EXPERIMENT III

Margolis, Dubno, and Hunt (1981) found that, for narrow notch widths, signal detection performance was worse in band-reject noise than in white noise, and improved as the notch was filled in with additional noise. This finding is reminiscent of the results of Experiment II in which performance was worse when the tones to be discriminated were close to the sharp cutoff of the fixed low-pass noise than when they were presented in unfiltered noise, and DLs decreased when high-pass noise was added. Both cases had the tones presented close to sharp spectral edges, and the additional noise which led to improved performance could have served to "blur" the sharp edges. In Experiment III the effect of variations in the sharpness of spectral edges on frequency discrimination was investigated by varying the slopes of the skirts of high- and low-pass filters.

Procedure

Two of the listeners who participated in Experi-

ments I and II plus an additional two listeners with no previous experience with the frequency discrimination task served as subjects. The same basic adaptive forced-choice procedure as was used in the previous experiments was employed in Experiment III to estimate two points on the psychometric function—those yielding 70.7 and 79.4 percent correct responses. The mean of the frequency separations corresponding to these two values was taken as the DL.

There were two sets of conditions in the experiment. In the first set of conditions the standard frequency was 975 Hz, and the tones to be discriminated were presented in low-pass noise with a frequency cutoff of 1000 Hz and a filter skirt with a slope of either 96, 72, or 36 dB/octave. In the second set of conditions the corresponding high-pass noises were used and the standard frequency was 1025 Hz. (In each case there were also control conditions in which the tones were presented in the quiet, and in unfiltered white noise.) The spectral level of the noise in the passband was 35 dB SPL per cycle. All tones had a duration of 100 ms and an intensity of 65 dB SPL.

Results and Discussion

In both sets of conditions there was evidence of an edge effect. In low-pass noise with a slope of 96 dB/octave the mean DL was 4.10 Hz (in comparison, the mean DL in unfiltered white noise was 3.60 Hz). Decreasing the slope to 72 and 36 dB/octave resulted in DLs of 4.07 and 3.34 Hz, respectively, and the difference between the DLs obtained in the noises with 96 and 36 dB/octave slopes was statistically significant. Thus performance improved as the sharpness of the spectral edge decreased, even though more noise power was present with the lower slopes.

In the high-pass noise conditions the mean DLs for the filter slopes of 96, 72, and 36 dB/octave were 3.50, 3.01, and 3.37 Hz, respectively. (The mean DL in unfiltered white noise was 3.84 Hz.) The difference between the values for slopes of 96 and 72 dB/octave was statistically significant, but that between the values for slopes of 96 and 36 dB/octave was not. Thus performance first improved and then declined as the sharpness of the edge decreased.

The relatively poorer performance in low-pass noise than in high-pass noise, and the finding that performance improved with decreasing slope in low-pass noise, may be a reflection both of an impairment which declines as the sharpness of the edge decreases and of the fact that (as observed earlier) low-pass noise is inherently more disruptive of performance than high-pass noise. Similar reasoning may also account for the non-monotonic changes in performance observed in the high-pass noise conditions.

REFERENCES


ON THE DISCRIMINATION OF DEVIATION OF THE PITCH OF TONE BURSTS.

Yasuhiro Yamashita and Masayuki Matsui
Department of Architecture, Faculty of Engineering,
Shinshu University
500 Wakasato Nagano Japan

1. INTRODUCTION

The sound in music and conversation, furthermore, in our surrounding almost change in time and in sound pressure. The real sounds involve the instantaneous changing sound pressure and frequency.

The purpose of the present work, therefore, was to measure the discrimination of the pitch of tone bursts changed frequency. In this study, the frequency change is defined as the difference between initial and final frequency. The change is linear in time, sometimes rising and sometimes falling.

On account of space consideration, the report was limited to measure the results of discrimination in case of rising frequency change.

2. GENERAL PROCEDURE

A. Apparatus

The bursts used throughout the experiments were produced by apparatus diagrammed in Fig. 1. The apparatus is able to generate two of tone bursts in which frequency changed linearly and separately from an initial value ($f_1$) to a final one ($f_2$) during the burst duration. Furthermore, the apparatus is able to change the burst duration separately.

Main circuits of the apparatus were three pulse generators and two shaping circuits. Two of pulses were for the duration of comparison for the interval between comparison burst and standard one. From shaping circuit, two of the stepped rising voltage waves were made separately from the pulse that was divided into 32 parts. Here the total duration of pulse was fixed up to the pulse generator circuit, i.e. the shaping circuit produces the voltage controlling the frequency of voltage-controlled oscillator (VCG).

![Block Diagram](image)

Fig. 1 The block diagram shows components for providing tone bursts with changing frequency. P1, P2, P3 are pulse generator, VCG is voltage controlled oscillator, A & B is voltage function generator circuit, ATT is attenuator of gain, MIX is mixer of A & B and W is window circuit in part of rising frequency change and of falling one.

In this way stimuli signals were generated by passing the voltage waves and the pulse for the duration of bursts through the window of switching pulse circuit. But as these stimuli signals have rasping click, these signals were shaping by passing through the window of 5 ms time constant for the part of rising and falling frequency change under synchronized switching pulse.

Both bursts were presented to one ear through the condenser type earphone. Schematic diagram of stimulus presentation are shown in Fig. 2.

These two stimuli (A, B) are for the comparison burst and for the test one. The duration of burst A and B was 0.4 sec and the pauses between the bursts were 1.6 sec. The variable dimension of the test burst within each series of all experiments was changed in a random order. Each test burst was matched 12 times by each subject.

B. Testing procedure

Three subjects (female student) about 22 years old, with normal audiograms, were used. Their task was to match, by the method of adjustment, the pitch of comparison tone burst to that of the test burst.

The SPL in earphone was controlled constant throughout the bursts at 75 dB re. 0.0002 pascal by calibrating in artificial 6 cc ear coupler.

3. RESULTS

A. Discrimination of the rising frequency change as functions of tone burst duration.

The results of discrimination of frequency change as functions of burst duration are shown in Fig. 3. When the frequency change of standard stimulus was in 300 Hz and when an initial frequency was set the value of 1000 Hz. The duration of both stimuli were changed from the value of 25 ms to 900 ms.

The means and standard deviations of the judgment of three subjects and are plotted as functions of burst duration in Fig. 3. The curves show that the discrimination of frequency change are nearly constant in disregard of the change for the burst duration.

In case of the pure tone bursts, it has been shown that the change for the burst durations became more short the more discrimination of frequency change became large, i.e. it has understood that an actual feeling for the pitch of tone burst is lost. But the present results do not show a tendency to the bursts of rising frequency change.

![Diagram](image)

Fig. 2 Presentation of stimulus signals. Stimulus A is the standard signal and stimulus B is the test one.
B. Discrimination for the pitch of rising frequency change of tone burst.

Tone bursts in which frequency changed linearly from an initial value ($f_i$) to a final one ($f_f$) were used in case of $f_i$ was fixed or in case of $f_f$ was fixed. The duration of stimulus A and B was fixed 0.2 sec and the interval of presentation on every stimulus was fixed 0.4 sec. The pauses between the pair of bursts were fixed 1.6 sec.

The initial frequency of standard stimuli were determined the value of 500 Hz, 1000 Hz, 2000 Hz and the rating of frequency change were determined the value of 1/20, 1/5, 1/2, 1/1.

Above the parameter of the bursts used in these experiments are shown in Table.1. The results of discrimination of frequency change in case of $f_i$ was fixed and the results in case of $f_f$ are shown together in Fig.4.

Generally the results in case of $f_i$ (show left side in graph) are shown larger than the results in case of $f_f$ (show right side). These may be understood that an actual feeling for the pitch of tone bursts will be inclined to the final frequency ($f_f$) and that the discrimination of the initial frequency ($f_i$) are not good.

Table.1 Parameter of stimuli used in experiment.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>$f_i$</th>
<th>$f_f$</th>
<th>1/20</th>
<th>1/5</th>
<th>1/2</th>
<th>1/1</th>
</tr>
</thead>
<tbody>
<tr>
<td>500Hz</td>
<td>500Hz</td>
<td>500Hz</td>
<td>500Hz</td>
<td>500Hz</td>
<td>500Hz</td>
<td>500Hz</td>
</tr>
<tr>
<td>1000Hz</td>
<td>1500Hz</td>
<td>1500Hz</td>
<td>1500Hz</td>
<td>1500Hz</td>
<td>1500Hz</td>
<td>1500Hz</td>
</tr>
<tr>
<td>2000Hz</td>
<td>2500Hz</td>
<td>2500Hz</td>
<td>2500Hz</td>
<td>2500Hz</td>
<td>2500Hz</td>
<td>2500Hz</td>
</tr>
<tr>
<td>4000Hz</td>
<td>5000Hz</td>
<td>5000Hz</td>
<td>5000Hz</td>
<td>5000Hz</td>
<td>5000Hz</td>
<td>5000Hz</td>
</tr>
</tbody>
</table>

ACKNOWLEDGMENTS

We wish to express our appreciation to Tadashi Takeuchi of this University who take part in many discussion concerning this work and contributed many suggestions. And we wish to thank Toshi Ido for his able assistance with the experiments.

REFERENCES

1) Turnbull, W.W. "Pitch Discrimination as a Function of Tonal Duration", J. Exp. Psychol., 34, 302-316 (1944)
INTENSITY DISCRIMINATION FOR SOUNDS WHOSE INTENSITIES CONTINUOUSLY CHANGE WITH TIME

Toshiie Nakamura
College of General Education, Osaka University
Machikaneyama-cho, Toyonaka, Osaka, 560 Japan

INTRODUCTION

In our constantly changing environment we are exposed to sounds of varying intensity and duration such as speech, music and traffic noise. Our recognition of acoustical information is based on discriminating and following fluctuating sounds over long periods of time. In order to develop a clear view of the hearing mechanism, it is important to examine to what extent the ear is able to discriminate and follow random temporal changes of sound pressure level (SPL). Research concerning the differential limen (DL) of auditory stimuli has typically been carried out with stimuli which are at some constant value or of sinusoidal intensity variation. Only a few studies have used continuously changing tonal intensities to examine loudness DL, and in these the changing tones are presented discretely. In reality, rather than turning up abruptly, stimuli are often continuously present with random changes during long periods in our daily lives. Consequently, it is important to increase our knowledge of the auditory system's capacity to discriminate continuously changing stimuli in a sound stream. This experiment was designed to examine the nature of our discrimination of intensity changes in such a situation.

EXPERIMENT

Stimuli

A tone of 1000 Hz was used. The stimulus was of more than 10 min total duration. It varied in intensity within the range of 70 ± 5 dB SPL, and consisted of 47 kinds of intensity pattern with two variables: (1) increasing and decreasing rates of 2, 0.5, 0.2 and 0.1 dB per second, plus constant intensity tone with seven different durations; and (2) complete intensity change of 5, 2, 1, 0.5, or no change. Consequently, each pattern had its own duration, defined by these variables (see Table 1). The 47 kinds of intensity pattern were presented successively in random order, connected by a steady intensity tone. An example of a stimulus is shown in Fig.1.

Table 1 Condition of stimulus

<table>
<thead>
<tr>
<th>Rate of change (dB/sec)</th>
<th>Amount of increment or decrement (dB) and duration (sec) in the parentheses</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5(2.5), 2(1), 1(0.5) and 0.5(0.25)</td>
</tr>
<tr>
<td>1</td>
<td>5(5), 2(2), 1(1) and 0.5(0.5)</td>
</tr>
<tr>
<td>0.5</td>
<td>5(10), 2(4), 1(2) and 0.5(1)</td>
</tr>
<tr>
<td>0.2</td>
<td>5(25), 2(10), 1(5) and 0.5(2.5)</td>
</tr>
<tr>
<td>0.1</td>
<td>5(50), 2(20), 1(10) and 0.5(5)</td>
</tr>
<tr>
<td>0</td>
<td>5(60), 2(20), 1(10), 5, 2, 1 and 0.5</td>
</tr>
</tbody>
</table>

SPL in dB

75 | 70 | 65 |
---|---|---|
5 sec | 12 min | Light on |

Fig.1 Example of stimulus

Procedure

The subject was seated in a soundproof room with the earphone on, monaural listening being used. A method developed by the author for rating the subjective impressions for parts of a long-term sound was used. With this method, the subject was given a visual clue simultaneously with the tone, in the form of an LED (light-emitting diode) signalling the subject to judge the loudness change of the stimulus during the LED's lit period. The extinguishing of the LED signalled the subject to indicate his judgment by pressing one of 3 response buttons labeled "getting softer", "no change" and "getting louder"; in addition, the subject was allowed to push the button "impossible" when he could not grasp the pattern during the LED's lit period. (Responses of impossible were only one per cent of all responses and concentrated in the patterns with the shortest duration).

Each subject was presented with three stimuli, consisting of the same 47 change patterns but in a different order. The procedure for presenting stimulus was as follows: when the subject moved the starting switch, the stimulus began at 70 dB level, continued for 7 sec and then, without a break, one of the 47 patterns appeared, accompanied by the LED. Though the LED turned off at the end of a given pattern and signalled the subject to give a response, the tone remained at the last level of the pattern until the next pattern began at that SPL, triggered by the subject's key-pressing. In this way, the subject was exposed to the 47 patterns of tone successively at his own pace. After practice for the task, there were three settings for this program for each subject.

Apparatus

Fig.2 is a block diagram of the apparatus used in the experiment. A pure tone generated by a function generator (IWATSU FG-350) was fed into a sound control system made up of a computer (SHARP MS-80B), a voltage control attenuator and a digital voltage generator. Using this system for controlling intensity change (0.1 dB step) and duration (1 msec step), it was possible to construct a changing stimulus during the experiment and present a visual signal by means of LED. Besides, the subject's responses were fed into the computer and stored on discs. The stimulus was smoothed by a 1/3 octave bandpass filter (NF-E-3201AS). The stimulus then passed through an amplifier (YAMAHA CA-1000EII) to the earphone (Telephonics TDH-49).

Subjects

Twelve males and a female participated. All had normal hearing and none of them had previously participated in an auditory experiment.

RESULTS AND DISCUSSION

The results for all 13 subjects were grouped together, giving a total of 39 trials for any one pattern of intensity change. A DL for increasing and decreasing loudness was calculated (Table 2). It might seem from these data that differential sensitivity to continuous variation of sound pressure is poorer than is shown in conventional discrimination tests. However, it was found in other experiments by the present writer that with an extremely fast rate of intensity change, even if...

---

Fig.2 Block diagram of the apparatus
accompanied by continuously varying tone, there was small DL. Conventional discrimination tests using two separate tones presented successively gave poorer results than the present method if a long time interval was inserted between the two tones corresponding to the duration of a gradually changing stimulus. In short, it seems that the influence of the time-factor (the period required for stimulus presentation) on DL should be considered.

These findings are summarized in Table 3. The experiments were designed to compare three methods of stimulus-presentation in DL measurement: (1) changing stimulus as part of tonal stream, (2) changing stimulus presented in isolation, and (3) two separate steady tones, in the manner of traditional DL measurement except that the time interval between the two tones was not constant but varied from 0 sec to 15 sec, corresponding to the duration of changing stimulus in (1) and (2). These data suggest that the ability to discriminate is better with a stimulus in a continuously changing stream than with an isolated stimulus. The latter situation is far different from experience in daily life. It should be noticed that the smaller DLs in Table 3(1) than in Table 2 probably were due to the greater number of cues for discrimination in the former; that is, where a stimulus contained the 2-sec steady intensities at the start and finish. Respecting loudness and/or noisiness estimation, the present author’s previous study showed that judgments for a stimulus in a tonal stream were more consistent than for an isolated tone.

In Fig.3 the proportion of the responses “getting softer” and “getting louder” is plotted as a function of the total extent of SPL change. When the SPL change is 5 dB (decreasing and increasing) the subjects’ performances are completely correct at every rate of change except the slowest rate of 0.1 dB/sec. 2 dB change is given almost 100% correct judgments only at a rate of 2 dB/sec; it is given fewer correct judgments as the rate of change becomes slower. Apparently the subject tends to respond to the rate. We should notice, however, that the duration of each pattern is defined by the rate of change. That is, the duration increased as the rate of change became slower, when the total change of intensity change was the same.

In order to clarify the influence of duration on the judgment of loudness in this experiment, Fig.4 shows the results for steady tone. The results indicate that as the duration of the steady tone becomes longer the percentage of correct judgments, “no change”, decreases, and subjects are biased toward the “getting louder” response. That is, even with no change in SPL they respond with “getting louder” more often than with “getting softer.” The stronger tendency to respond “getting louder” has been examined by the present author in the case when the stimuli were presented in isolation. Findings are shown in Fig.5. Comparison between Fig.4 and Fig.5 reveals differences between auditory perception with a stream of sound and with isolated sound. The tendency to respond “getting louder” can be seen in Fig.3, where a stimulus of increasing SPL is correctly judged more often than of decreasing SPL, especially at the slowest rate of change, 0.1 dB/sec, i.e., with the longest durations. This finding is of interest with respect to the issue of loudness adaptation. If adaptation occurs, it would be expected that the response to a constant SPL would be “getting softer”, and a stimulus whose SPL is decreasing should be correctly identified more often than one which is increasing. But the present data run counter to the adaptation predictions mentioned above. This trend counter to loudness adaptation has also been observed in the present author’s earlier study on listening music. Although the influence of the rate of change on DL has some significance, duration effect must be taken into account, since it makes discrimination of increasing SPL easy and that of decreasing SPL difficult.

References
GENERATION OF BINAURAL SIGNALS FOR RESEARCH AND HOME ENTERTAINMENT

C. Pössel, J. Schröter, M. Opitz, P.L. Divenyi and J. Blauert

Lehrstuhl f. allgemeine Elektrotechnik und Akustik
Ruhr-Universität Bochum, PO Box 102148
D-44780 Bochum, F.R. Germany

Abstract

A convenient way of generating head-related binaural acoustic signals consists of convolving nonreverberant sound signals with binaural impulse responses of the external ears. The method of obtaining these impulse responses will be described. Two applications will be presented.

Measurement of Transfer Functions (TFs) of the External Ear

Computer aided measurements have been carried out both on a dummy head and on a set of human subjects with a measuring set-up as sketched in Fig. 1.

![Computer aided digital TF-measurement](image)

Fig. 1 Computer aided digital TF-measurement

Fig. 2 shows the set of angles of sound incidence selected within the upper hemisphere. The elevation angle \( \theta \) was varied in steps of 10° in the range of \(-10^\circ\) - \(90^\circ\) by using a so called "audiophonic" consisting of 11 loudspeakers mounted along a vertically suspended circle segment (\( R = 2.50m \)). The azimuth angle \( \phi \) could be adjusted in steps of 90° by using a revolving chair in the center of the audiophonic. Exact head positioning was achieved by using three horizontal laser beams which were focused into the center of the audiophonic. The heads were positioned for the beam to aim at each ear canal and the tip of the nose.

The signals were picked up by miniature microphones within each ear canal. Reference measurements of headphone TFs had previously been made with the same microphone position (blocked meatus, 5 mm insert depth) for later binaural reproduction through these headphones. Finally all free-field functions were divided by the previously measured TFs of the loudspeakers in use to obtain the pure TF of the external ear.

Analysis of Measured TFs

Digital filtering of "dry" signals with a binaural external ear filter can be accomplished either in the frequency domain (overlap-add technique) or in the time domain (FIR filter). FIR-filter coefficients are provided directly by the external ear impulse response. Bearing in mind that the measured data were to be utilized for real-time digital-convolution purposes, first of all, we were interested in the duration of binaural impulse responses.

The measurement procedure itself was carried out in the frequency domain. Not-measured frequency range between DC and \( f_0 = 117 \) kHz was estimated by a linear extrapolation, whereas the gap between \( f_0 \) = 16 kHz and the Nyquist frequency was filled by estimating a "tail" \( \frac{1}{f} \) before calculating the appropriate external ear impulse response (see Fig. 5). This was necessary in order to eliminate the influence of the steep slope at \( f \) upon the time domain (ripples on impulse response), and to avoid loss of low frequency components by the filtering process.

In order to check the required time-window length for binaural impulse responses, and to get an impression about the energy distribution over the time axis, we plotted the signal envelope of the external ear impulse response at 360°-rotation in the horizontal plane (Fig. 3).

![Signal envelope of external-ear impulse response](image)

Fig. 3 Signal envelope of external-ear impulse response, dummy head, right ear

The functions were measured at 90° intervals with a dummy head equipped with a set of a typical pinna (equal NEUMANN KU 811). "Signal envelope" means the magnitude of the analytic signal which had been calculated by a Hilbert transformation \( \frac{1}{f} \). The maximum delay between ipsi- and contralateral side is 500 µs, and the main energy of each single response is distributed over about 500 µs.
Fig. 4 illustrates the smoothing effect which windowing in the time domain has on the frequency response of the "Front" function (curves are 5 db spaced by plot, Hamming window centered at signal maximum). Smoothing grows with decreasing window length. Linear distortion does not exceed +/- 2 db, at least for 64 points of window length, whereas the 32-point window obviously provides too much distortion.

Without mentioning the psychoacoustical evidence of these windowing procedures one may yet conclude from Fig.3 and Fig. 4 that a time-window length of 5 ms (equivalent to 200 points here) may be sufficient for the storage of binaural impulse responses.

In order to analyse the monaural spectral and temporal structures of the external-ear TFs, the median plane was examined in more detail in the following figures:

Fig.5 TF of the external ear, mean values of 11 subjects, median plane

Fig.5 reveals a significant up- and downwards shift of the 1 kHz notch and, to some minor but still significant extent, an alteration of the midrange frequency response, all of which correspond to the psychoacoustically proved "directional bands" /2/.

We found that external-ear TFs are not minimum-phase functions, but contain significant allpass-group-delay components at different frequencies; e.g. the allpass-group-delay spectra of the median-plane TFs contain at least two sharp, major peaks at different frequencies for the "front" and "rear" TF, whereas the "top" function has almost minimum-phase character.

In Fig. 6 the signal envelopes of the mean impulse responses of the median plane are compared to the envelopes of the corresponding minimum-phase functions. The removal of the allpass components leads to a certain energy concentration in each curve and "orders" its peaks. Nevertheless, the resulting envelopes reveal a very clear pattern of three peaks for the front, one peak for the top and two peaks for the rear, which is in agreement with the data of /3/.

Application to a Binaural Mixing Console

By using the described window technique a filter set of binaural impulse responses for the upper hemisphere was derived from the measured transfer functions. Gaps due to non-measured angles of sound incidence could be filled imperceptibly by interpolating neighboring impulse responses - as listening tests with simulated "walk-arounds" have shown. Simulations of free-field listening with stationary and moving sound sources are perceived as natural; even binaural room simulations by means of a binaural mirror-image model have been conducted successfully in first experiments.

Application to a Clinical Test of Spatial Hearing

The same method has been applied to simulation of free-field listening through headphones to investigate hearing abilities of patients with unilateral lesions of the anterior temporal lobe (ATL) /4/.

With the method presented in this paper, multiple test signals can be economically computed, saving large-scale anechoic chambers and expensive hardware.

Literature


NEW APPROACH TO THE RISE TIME DIFFERENTIAL SENSITIVITY

Jacek Smurzyński

Warsaw Academy of Music, 2 Okólnik, 00-368 Warsaw, Poland

INTRODUCTION

Within the last five years there has been a new interest in categorical-like perception of nonspeech signals. New investigations argued that categorical perception was questionable for both speech and nonspeech continua. One of the criteria of categorical perception says that subjects are able to discriminate only sounds that are identified as belonging to different categories. To examine if this criterion holds it should be known how large the differences in onset values of acoustic stimuli have to be in order to be perceptually different. Yet, surprisingly little is known about the difference limens (DEL) of the rise time. Van Heuven and van den Broecke /1979/ established DELs for rise time of 1000 Hz sine waves as well as white noise bursts by means of an adjustment method. The Weber fraction was about 25%. Similar results were obtained by fewley - Port and Pisoni /1984/ in the adaptive tracking procedure for 300 Hz sawtooth waves. In the present experiment new DEL values were obtained for sine waves by a constant stimulus method with multiple comparison in a single trial.

METHOD

Stimuli

Amplitude envelopes of stimuli with linear onset envelopes were generated digitally on an Olivetti M-20 microcomputer. The steady-state portion was followed by a 40 msec linear decay. The overall length of the stimuli was 256 msec. The amplitude envelopes were output under computer control at a 4 kHz sampling rate through a 10 bit multiplexing D/A converter and multiplied by 1000 Hz or a 333 Hz sine wave /Hewlett-Packard 3312 function generator/. The stimuli were then presented binaurally through dynamic headphones WH-60 at 70 dB above threshold in a quiet testing room. Each stimulus started at a positive-going zero crossing. Rise time duration at the headphones was carefully measured for each of the stimuli using an artificial ear. The shape of the amplitude envelope at the stimulus on-set was clearly linear.

Procedure

Six basic stimuli with onset time of 10, 20, 30, 40, 50 and 60 msec were used in six blocks of 24 trials for both frequencies. Each trial consisted of a pair of pulses: a standard - base stimulus and a comparison one. The onset of the comparison was equal to the standard or longer by 0.5, 1.0, 1.5 or 2.0 msec. A subject used the remote control to call amplitude envelopes of the standard and the comparison stored in the computer memory. Each stimulus lasting 256 msec was played back every one second. A subject could call both stimuli in a single trial as many times as he needed to judge a pair as the same or different. There were no restrictions on the number or order of the callings on the part of the subjects. The answer terminated a trial. An extra five sec interval between trials.

In a block of 24 trials 12 comparison stimuli were equal to the standard. In the other 12 trials each of the four different values of the standard were presented three times. The order of presentation was randomized. Five sessions of twenty four trials gave 15 discriminations per stimulus per subject per frequency.

SUBJECTS

Three subjects, two males and one female, graduate students at the Sound Engineering Department of the Academy of Music, participated in the experiments. All subjects were aware of the purpose of the test and were paid for their participation. All of them passed a special course of so-called "timbre solfeggio". The goal of this course is to develop timbre perception skills that is, to increase timbre sensitivity and improve timbre memory /Katowski, 1985/.

RESULTS

Basing on subject's answers psychometric curves for all twelve experimental conditions were found. The value of the rise time corresponding to 75% of discrimination was taken as a measure of the DEL.

The values of DEL for three subjects, six standards and two frequencies are presented in the table.

<table>
<thead>
<tr>
<th>Frequency 1000 Hz</th>
<th>DEL</th>
<th>Subject: TR</th>
<th>DEL</th>
<th>Subject: MW</th>
<th>DEL</th>
<th>Subject: LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.3</td>
<td>1.2</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.4</td>
<td>1.0</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>1.8</td>
<td>1.4</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>1.3</td>
<td>1.7</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.2</td>
<td>1.6</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>1.8</td>
<td>0.8</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DISCUSSION

Frequency of the signal seems to have no influence on the threshold, as DELs for 1000 Hz and 333 Hz are very similar.

The Weber fraction derived from the data fall in 2-13% range. These values are much smaller than those obtained by Fewley-Port and Pisoni /1984/ and van Heuven and van den Broecke /1979/. There are at least two parameters having an effect on the results. The first one is the method. A subject could compare both stimuli from a
Frequency 333 Hz

<table>
<thead>
<tr>
<th>Standard rise time msec</th>
<th>Difference limen msec Subject:TW</th>
<th>Difference limen msec Subject:MW</th>
<th>Difference limen msec Subject:TK</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>20</td>
<td>1.5</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>30</td>
<td>1.4</td>
<td>1.3</td>
<td>0.8</td>
</tr>
<tr>
<td>40</td>
<td>1.3</td>
<td>0.9</td>
<td>1.2</td>
</tr>
<tr>
<td>50</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>60</td>
<td>1.4</td>
<td>1.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>

trial as many times as he wanted and in any order being the best one for him, to judge the rise times. Thus, each judgment was based on multiple comparison and that could decrease the DL. The effect of the test procedure on the DL for another attributes of sound is wide known /see for example Łetowski, 1982 for frequency differentiation/. The second reason is connected with the subjects who are highly experienced in timbre assessment. One part of the program of timbre solfeggio deals with the discrimination and memorizing typical rise and decay times of various stimuli.

A new phenomenon when comparing with previous results is, that received DLs do not follow Weber's law, i.e., the values of DL are rather constant for the rise time from 10-60 msec range. It seems that subjects perceived changes of the rise time absolutely rather than relatively. Generally, the increase of the rise time by 0.8 – 1.0 msec was enough to reach the differential threshold. This value is much smaller than that needed to confirm the existence of categorical-like perception. A preliminary experiment of absolute identification of the rise time gave the accuracy of about ± 10 msec /Smurzyński, 1986/. That means, subjects were able to discriminate the rise time much more precisely than to identify them. Thus, one main criterion of categorical perception was not fulfilled.

REFERENCES

4. Łetowski, T., Development of technical listening skills. Timbre solfeggio. Journal of the Audio Engineering Society, 1985, 33, 240-244
5. Smurzyński, J., Noncategorical identifica-
THE EFFECT OF AMPLITUDE ENVELOPE ON THE PITCH OF SHORT SINE-WAVE AND COMPLEX TONES

Thomas D. Rossing and Adrianus J. M. Houtsmul

Department of Physics, Northern Illinois University, DeKalb, IL 60115, USA and
Institute for Perception Research, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

Although the pitch of a sine-wave tone is mainly determined by its frequency, it may depend also on other parameters such as intensity, duration, envelope shape, masking tones or noise, or a preceding or following tone. Pitch shifts due to changes in these parameters vary from subject to subject, and they are not easily predicted on the basis of present models of hearing. The purpose of these studies was to investigate, in a systematic way, the dependence of pitch on amplitude envelope.

The perceived pitch of a short exponentially falling sine-wave tone is judged higher than that of a gated sine-wave tone of the same frequency and energy (Hartmann, 1978). Hartmann's experiment was carried out at one intensity (89 dB for the gated tone) and one decay rate (1 dB/ms). In the present study, pitch matches were made using both rising and falling exponential envelopes, rates of rise or fall from 0.5 to 8 dB/ms, and reference intensities ranging from 70 to 100 dB. At each reference intensity \( I_{\text{ref}} \) and each decay rate \( x \), each subject matched 80 pairs of tones within the test frequency range between 200 and 3200 Hz. The tone with the exponentially rising or falling envelope alternated with a 40 ms gated comparison tone. Initial amplitudes were selected so that both tones had equal energy, and silent interval of 860 ms occurred between tones. The subject adjusted the frequency of the comparison tone until both tones appeared to have the same pitch. Test tone frequencies were chosen randomly within each octave band.

Results obtained for one reference intensity are shown in Fig. 1a and for one rate of exponential fall in Fig. 1b. In both cases, the frequency difference \( \Delta f \) for equal pitch increases logarithmically with frequency. All the data for falling exponential envelope can be described by the relationship

\[
\Delta f = 0.0221 e^{0.07 I} (\log x + 0.458)(\log f + 2)
\]

where \( I \) is the sound pressure level of the comparison tone in \( \text{dB} \), \( x \) is the decay rate in \( \text{dB/ms} \), and \( f \) is the test tone frequency in Hz. This empirical formula, which describes the data as a product of three functions, is not intended to provide insight into a causal relationship between pitch, intensity and decay rate.

Further experiments indicate that exponentially rising envelopes induce comparable upward pitch shifts. This finding eliminates models that assume some kind of causal relationship between instantaneous signal amplitude and periodicity in neural firings, for these models would predict opposite shifts using rising and falling exponential envelopes.

![Fig. 1. Pitch shift as a function of frequency for various exponential signal decay rates (a) and for different intensities (b). Test tone frequency is plotted in octaves re 200 Hz.](image)

The results of these experiments, plus some additional experiments that investigated the dependence of pitch on amplitude and duration, suggest that the envelope-induced pitch shift may not be caused by signal envelope per se, but rather by some envelope-influenced quantity such as average intensity. One such quantity is the time average sound pressure level averaged above an arbitrary reference level, say, 50 dB. Levels thus averaged correlate well with the observed values of \( \Delta f \) for both rising and falling exponential envelopes.

Complex tones

In another set of related experiments we investigated the pitch shift of complex tones consisting of two sinusoidal tones with frequencies \( f_1 = mf + 50 \) and \( f_2 = (m+1)f + 50 \), where \( m = 3, 4 \), or 5. Subjects were asked to judge the "low" pitch as well as the pitches of the individual components \( f_1 \) and \( f_2 \). The pitch-watching procedure was similar to that used in the sinusoidal tone experiments.

We find that shifts in \( f_1 \) are generally less, and shifts in \( f_2 \) are generally greater than the shifts observed in isolated sine-wave tones of comparable frequencies. Shifts in the perceived pitch of both components \( f_1 \) and \( f_2 \) show only slight dependence on the envelope decay rate in contrast to isolated sine-wave tones where \( \Delta f \) increases monotonically with exponential decay rate.
The "low" pitch of the complex tone is generally found to be 0-1.5% higher than that computed from the first order pitch shift using the linear estimator \( f = f_1/2m + f_2/(2m+1) \). The pitch shift appears to depend less on envelope decay rate than the pitch of sine-wave tones of comparable frequency.

Acknowledgements

The authors thank J. 't Hart and W. M. Wagenaars for serving as subjects and also J. R. Cohen, H. Duithuis, and W. M. Hartmann for helpful conversations. This work was supported by a grant from the Dutch Foundation for Pure Research (ZWO).

References

PERCEIVED PITCH OF TWO-COMPONENT FM-AM TONES

S. Iwamiya and T. Miyakura

Dept. of Acoustic Design, Kyushu Inst. of Design, Fukuoka, Minami-Ku, Fukuoka, 815 JAPAN

INTRODUCTION

There are many studies on pitch perception of steady sounds. However, the components of musical or speech sounds are generally accompanied with frequency modulation (FM) and amplitude modulation (AM), and pitch perception of fluctuating tones is not studied so extensively. The purpose of the present study is to make clear the effect of FM and AM of each component on pitch perception of complex tones.

We perceive steady and average pitch from FM tones (like vibra-tone) accompanied with instantaneous pitch fluctuation, and the pitch of periodically FM tones is perceived near their carrier frequencies (the middle of the extent of FM) (Shonle and Noran, 1960; Iwamiya et al., 1983). The pitch of one-component FM-AM tones is matched to the time average of instantaneous pitch fluctuation weighted with their loudness fluctuation (Iwamiya and Kitamura, 1983; Iwamiya et al., 1984; Iwamiya and Fujisawa, 1985). The weighted pitch averaging process is hypothesized to be in the auditory system. FM-AM tones are defined as fluctuating tones whose frequency and amplitude are modulated periodically. The existence of the weighted pitch averaging process is also suggested by Feth et al. (1982).

In the present study, when the carrier wave consists with two harmonic components, the perceived pitch of FM-AM tones is measured as a function of the phase difference between FM and AM of each component using a method of adjustment.

EXPERIMENTAL METHOD

Two-component FM-AM tones were generated by a computer and delivered to a 12-bit digital-to-analog converter at a sampling rate of 20 kHz. They were filtered by a low-pass filter with a cutoff frequency of 4 kHz and 96 dB/oct slope.

The carrier wave consisted of 1640 Hz and 880 Hz harmonic sinusoidal waves at equal amplitude. The modulation waveform of each component was sinusoidal and the modulation rate was 7 Hz. The phase between FM and AM of each component was set for in-phase or anti-phase. In the case of in-phase condition, frequency and amplitude simultaneously increase and decrease. Oppositely, in the case of anti-phase condition, frequency increases while amplitude decreases, and frequency decreases while amplitude increases. The phase between FM or AM of the fundamental component (1st FM or 1st AM) and FM or AM of the second component (2nd FM or 2nd AM) was also set for in-phase or anti-phase.

The extent of FM of each component (E) is defined as 12000log2 (the highest frequency/the lowest frequency) cents. The degree of AM (AM) is (maximum amplitude–minimum amplitude)/(maximum amplitude + minimum amplitude).

The experiment was done under the three modulation conditions. In condition 1, E=50 cents and AM=1.0 for both the components. In condition 2, E=50 cents and AM=0.0 (FM only) or E=0 cent and AM=1.0 (AM only) for the 2nd component. In condition 3, E=50 cents and AM=0.0 or E=0 cent and AM=1.0 for the fundamental; E=50 cents and AM=1.0 for the 2nd component.

The phases between FM and AM of each component and those between 1st FM (AM) and 2nd FM (AM) of the sound stimuli under each condition are represented in Figs. 1 to 3.

The task of the subjects was to adjust the pitch of the pure tone to match that of the two-component FM-AM tone. They could listen to both the signals alternatively as long as they desired. The signals were presented dichotically by a headphone (Stax SR-X/MK-3) at same loudness. The pure tones had a level of 70dB SPL. Five subjects with normal hearing, aged 21 to 25 years, participated. They did 10-time matching trials for each stimulus.

RESULTS

A setting frequency (Hz) for each trial is converted into a interval from the fundamental frequency (440 Hz) of the carrier wave in cents. The average settings and standard deviations for all the data are represented in Figs. 1 to 3.

Figure 1 shows that the pitch of two-component FM-AM tones 1 to 4 is higher than that of tones 5 to 8 under condition 1. The phase between 1st FM and 1st AM of tones 1 to 4 is in-phase and that of tones 5 to 8 is anti-phase. The pitch of two-component FM-AM tones at in-phase relation between FM and AM is shifted higher than that at anti-phase relation (Iwamiya et al., 1984). Therefore the pitch shift of two-component FM-AM tones is mainly determined by that of the fundamental FM-AM tones.

Moreover the pitch shift interval based on the phase difference between 1st FM and 1st AM varies as a function of the phase difference between 1st FM and 2nd AM and that between 1st FM and 2nd AM.

The pitch shift interval between tones 2 and 8 is larger than that between tones 1 and 7, and that between tones 4 and 6 is also slightly larger than that between tones 3 and 5. In tones 2, 4, 6, and 8, 1st FM and 2nd AM are anti-phase, and they are in-phase in the others. Therefore the pitch shift interval based on the phase difference between 1st FM and 1st AM is larger when 1st FM and 2nd AM are anti-phase than when they are in-phase under the same phase relation between 1st FM and 2nd AM.

The pitch shift interval is affected additively by the phase difference between 1st FM and 2nd AM and that between 1st FM and 2nd AM. The effect of the phase difference between 2nd FM and 2nd AM on the perceived pitch is not observed.

Figure 2 shows that the pitch shift is also mainly determined by the phase difference between 1st FM and 1st AM under condition 2. Furthermore, as the pitch shift interval between tones 9 and 11 is larger than that between tones 10 and 12, the effect of the phase difference between 1st FM and 2nd AM is confirmed. Similarly, the effect of the phase difference between 1st FM and 2nd AM is confirmed as the pitch shift interval between tones 14 and 15 is larger than that between tones 13 and 15.

Under condition 3, as the fundamentals of the sound stimuli are accompanied with either of FM or AM, the pitch shift based on the phase difference between 1st FM and 1st AM cannot be observed. In fact the pitch shifts of the sound stimuli in Figs. 3 are much smaller than those in Figs. 1 and 2. Therefore the dominance of the effect of the phase difference between 1st FM and 1st AM on the perceived
The pitch of steady complex tones is shifted when one of the partials is mistuned from harmonic relationship (Moore et al. 1985). The pitch shift by the mistuned partial is dominant within the lowest six harmonics of the twelve harmonics.

In the case of complex FM-AM tones, the pitch of each FM-AM component is shifted as a function of its phase difference between FM and AM when it is perceived solely. However, the perceived pitch of two-component FM-AM tones is almost determined by the pitch shift of the fundamental component. It is not affected by the pitch shift of the second component. Therefore the weighted pitch averaging process for pitch perception of FM-AM tones is considered to be after so-called "a central pattern recognizer" for residue pitch perception of complex tones. If this process was before "a central pattern recogniser", the pitch of two-component FM-AM tones would be affected by the pitch shift of both the components. This process functions on the basis of information of 1st FM and 1st AM.

The effect of the phase difference between 1st FM and 2nd FM (or 2nd FM) is hypothesized to be caused by the interaction between the auditory processing channels for modulations of each component. The function of the processing channel for 1st FM is activated by the interaction between the channels for 1st FM and 2nd FM when 1st FM and 2nd FM are anti-phase or suppressed when they are in-phase. Thus the perceived extent of 1st FM is larger when 1st FM and 2nd FM are anti-phase than when they are in-phase. As the pitch shift interval of FM-AM tones is proportional to the extent of FM (Iwamiya et al., 1984), it is larger when 1st FM and 2nd FM are anti-phase than when they are in-phase. Similarly, the pitch shift interval is larger when 1st FM and 2nd FM are anti-phase than when they are in-phase by the interaction of the processing channels for them. The effects of these interactions are additive.

REFERENCES


STUDIES ON THE PERCEPTION OF DISTORTION IN THE LIGHT OF AN AUDITORY MODEL

M.A. Karjalainen
Helsinki University of Technology, Acoustics Lab, Otakaari 5 A, 02150 Espoo, Finland

This paper describes experiments with a computational auditory model to study the perception of distortion in speech and audio signals. It is shown that the distortion measures based on the model fit well to the results of subjective tests. Some basic rules known from the psychoacoustic theory are shown to be valid for complex distortion concepts. A practical measurement system based on the results and utilizing a signal processor is under development.

INTRODUCTION

Auditory modelling by computational means makes it possible to study complex auditory phenomena by describing them quantitatively in the form of a model. For this purpose, a detailed model for short-time auditory spectrum analysis was constructed including the most important properties of hearing known from the psychoacoustic theory. Such properties are:

* Frequency sensitivity to match the equal loudness curve (60 dB).
* Frequency selectivity of 1 Bark (1 critical band) and masking effect in frequency domain.
* Temporal integration (time constant 0.1 - 0.2 s) and backward and forward masking effects.

COMPUTATIONAL MODEL FOR AUDITORY SPECTRUM

Auditory spectrum instead of Fourier spectrum, is the representation of signals gained by the auditory models. A filter-bank type of model was adopted in this study because of its analogy to the peripheral hearing mechanisms: basilar membrane and hair cells which can be considered as a multi-channel analyzer. It is sufficient to have 1 - 4 channels per one Bark in a computational model, which means 24 - 96 channels covering the 24 Bark audio range. The model used in this study had 48 channels with 0.5 Bark spacing which seemed to be a practical compromise between good resolution of spectral representation and low amount of computation.

Auditory filter bank

Fig. 1 illustrates one of the 48 channels of the filter-bank model. Each channel consists of a bandpass filter, a square-law rectifier, a fast linear and a slower nonlinear lowpass filter, and a dB-scaling output stage.

![Diagram of auditory filter bank](image)

The bandpass stages with 1 Bark bandwidth were 256th-order FIR filters carefully designed to have a frequency response which is the mirror image of the spreading function. This gives a good approximation of the desired masking properties in the frequency domain. The frequency sensitivity according to the equal loudness curve is easily obtained by proper selection of the gain coefficients in the filter-bank channels.

The next step in each channel of the model is to rectify the bandpass-filtered signal. Instead of half-wave rectification a square-law rule is applied which facilitates the implementation of temporal integration. To simulate the threshold of hearing a constant level is added to the output of the square-law element.

Auditory time constants

There are two lowpass filters after the rectification stage in the auditory filter bank. The first filter has a time constant of about 3 ms and is not very essential providing only some smoothing of the fast amplitude variations in the pitch frequency range.

The second filter is more important. Its purpose is to implement many effects, such as temporal integration and pre- and postmasking. Temporal integration is realized simply by linear time-order lowpass filtering with a time constant of about 100 ms.

Postmasking (forward masking) is a complicated phenomenon and not easy to be matched well in a computational model. A nonlinear rule is used for the second filter in the case of masking. The overall response of the second lowpass is:

\[ X5(n) = K1 * X4(n) + (1 - K1) * X5(n-1) \]

where \( X4 \) and \( X5 \) are the input and output of the filter, \( K1 \) and \( K2 \) filter coefficients and \( n \) the discrete time variable. A good value for \( K1 \) was found to be 0.0005 and for \( K2 0.0007 \) with a sampling frequency of 20 kHz.

DISTORTION MEASURES BASED ON AUDITORY SPECTRUM DISTANCE

If some aspects of a sound are changed, the perceived difference is found to correlate well with the corresponding change in auditory spectrum. From the theory of psychoacoustics we know that a deviation of about 2 dB over any Bark-channel is the just noticeable difference (JND) found in listening tests ("2 dB -rule").

The basic hypothesis in this study was the assumption that the deviation in the auditory spectrum caused by nonlinear distortion has a good correlation to the perceived level of distortion. In an earlier study with a simple model for auditory spectrum distance different metrics of spectral distance were compared. The maximum deviation over the Bark scale and the Euclidean distance of auditory spectrum vectors yielded good results. In this study the maximum deviation over Bark and time was used because of its simplicity.

Perception threshold for nonlinear distortion

The "2 dB -rule" was tested by distorting three Finnish sounds: /a/, /i/ and /s/ (one speaker) with three nonlinear distortions: square-law, crossover and clipping. Duration of the distorted sound was the third variable factor. Three subjects determined the JND levels of distortion by directly comparing distorted and undistorted signals in an anechoic chamber. The corresponding maximal distances in auditory spectra were computed by using the model. The results are plotted in Fig. 2 as a function of duration of the distorted sound.

Figure 2 shows that in general the results follow the "2 dB -rule" well. The JND threshold corresponds to 1 - 2.5 dB of auditory spectrum distance for all sounds and distortions. If temporal integration is not included in the model the threshold curve will depend more radically on the duration of the distorted sound.

The validity of the "2 dB -rule" was shown also in
other cases. Any linear distortion (frequency response error) exceeding this level can be heard. The test signal may be harmonic or random, but also a nonharmonic complex of harmonic signals. Spectral distance can be used also as a measure of signal-to-noise ratio (SNR). Instead of the "2 dB rule" the JND-threshold corresponds now to about 0.5 dB maximum auditory distance.

Subjective distortion vs. auditory spectrum distance

Distortion is perceived in practice without any nondistorted reference. In the present study we found that instead of the "2 dB rule" the threshold varied between 1.5 and 1.5 dB in the case of no pure reference. Things become even more complicated when one is trying to model the perception of distortion over the JND-level. In spite of these principal irregularities a series of experiments was carried out to relate some subjective indexes to objective spectral distances. In the first case the subjective distortion index (SDI) was fixed on a scale from 0 to 10 so that no distortion was a 0 rating and point 5 was fixed by listening to a reference distortion (5 dB spectral distance). To relate the subjective and objective measures the resulting curves which are averages for three listeners are shown in Fig. 3a. Differences between individual listeners were of the order of ±1 step on the subjective scale.

In another case the subjective scale was given by the descriptive definitions:

1. No audible distortion.
2. The listener supposes to have heard something like distortion but is not sure.
3. Distortion is on the just noticeable threshold.
4. Distortion is always perceived when concentrating on listening.
5. Distortion can be easily heard as "soft" distortion.
6. Distortion is now disturbing.
7. Listener feels some discomfort because of distortion but the sounds are still easily recognized.
8. Distortion is increased to the level where some problems of correct recognition exist.
9. Recognition of the sounds is like guessing.
10. Recognition of the sounds is impossible.

The results of this study are shown in Fig. 3b. The cases 3a and 3b do not show large differences and in general the relation between subjective and objective scales is monotonic and fairly systematic. Differences between speech sounds exist but negligible variations among listeners (not indicated in the figure) showed potential stability of the approach.

All cases in Fig. 3b where the SDI is equal to or more than 6 were interpreted as "disturbing" by the definition. The level of disturbance is important in practical evaluation of audio systems. The average thresholds of disturbance for different sounds were 6.5 dB(i), 9.7 dB(u), 11.7 dB(s) and 9.3 dB total average. Distortion type showed less effect on the results than the type of sound.

As a consequence of the results from these studies a new series of experiments will be started to find versions of the auditory spectrum distance having more systematic correspondence to the subjective distortion levels. This means that the original auditory spectrum must be modified e.g. to emphasize its local properties of formant-like resonances.

PRACTICAL DISTORTION MEASUREMENT

Based on experimental studies and the auditory model the design of a practical measurement system for nonlinear distortion in speech transmission is in progress. The implementation will be based on the TMS320 (Texas Instruments) signal processor and the Apple Macintosh personal computer. A small database of speech samples and speech-like signals to be used as test signals will be stored on a floppy disk. The system will contain 16 bit A/D- and D/A-converters for full audio range.

The procedure for a typical measurement of sound quality would be the following: a short sweep-like test signal with a flat spectrum is used first to measure the frequency response of the system. The effect of nonlinear distortion is then measured with several samples of real speech as the maximum auditory spectrum deviation over the Bark and time scales. The effect of frequency response quality compensation is finally yield the auditory measure of nonlinear distortion. It is possible to apply the same method to the measurement of auditory signal-to-noise ratio.

ACKNOWLEDGEMENTS

The development of the auditory model was sponsored by the Academy of Finland and the application of the model to distortion measurement was financed by the Finnish PTT.

REFERENCES

SPECTRAL DOMINANCE FOR NOISE SIGNALS WITH MONOAURAL AND DICHOTIC COMB-SPECTRA

Johan Raatgever and Mark J. Bakker

Applied Physics Dept., Delft University of Technology
P.O. Box 5046, 2600 GA Delft, The Netherlands

INTRODUCTION

Spectral dominance in monaural pitch perception

At present, the pitch of complex tones is generally explained as a consequence of their internal spectral representations. Different pitch theories have been developed based on spectral pattern recognition of information from the individual harmonics of complex sounds. In this processing not all frequency components are equally important. Lower resolved harmonics are often reported to contribute more than higher harmonics. Thus, spectral weighting is assumed accordingly. Several theories further assume that some components, sometimes even one, are most important or dominant in the perception of the low pitch of a complex tone (e.g. Ritsma, 1967; Bilsen, 1977; Raatgever and Bilsen, 1977; Terhardt et al., 1982). Ritsma (1967) introduced the idea of spectral dominance to account for the pitch of anharmonic signals. He concluded that the components in a small frequency band around the 3rd-5th harmonic are dominant in determining the pitch of the complex. Flomp (1967) found the position of the dominant frequency region to depend on the fundamental frequency. This point has been confirmed by Terhardt et al. (1982) and further elaborated in their pitch predicting algorithm. Recently, Moore et al. (1985) reported considerable differences in the position of the dominant region for individual observers. For complex tones with fundamental frequencies from 100-600 Hz they sometimes found dominance of the 1st, 2nd and 3rd harmonics while for other subjects the 4th and 5th seemed to be dominant. Bilsen (1977) reported that for the class of monaural complex noise signals with a cosine-shaped power spectrum (MRB or cosine-noise) the dominant region tends to be constantly positioned around the 4th harmonic for fundamental frequencies up to 1000 Hz.

It has often been suggested that a competition exists between the (low) pitch due to the complex of harmonics and the (spectral) pitch from the fundamental component (e.g. Flom, 1967; Terhardt et al., 1982). These two pitches conflict most clearly for anharmonic tone complexes. The pitch from the lowest harmonic takes over for higher fundamental frequencies. This transition frequency is reported to be near 600 Hz (Flom, 1967) or even lower, Terhardt et al. (1987), however, predict a transition frequency around 1200 Hz in their model.

Spectral dominance in binaural processing

Bilsen and Raatgever (1973) reported for low-frequency transients that the fixed frequency region around 600 Hz is dominant for lateralization. In later studies (Raatgever, 1976, 1980), experimental evidence was provided that, in general, the 600 Hz region is dominant for the lateralization of the frequency signals due to interaural phase or time differences. This fact has been confirmed in several recent studies (e.g. Henning, 1983).

The frequency region around 600 Hz has been reported to be dominant in the perception of some dichotic pitch phenomena too (Bilsen, 1977). Dichotic pitch could be fully explained by assuming central spectral information due to binaural interaction. It has been argued that the supposed central spectra of dichotic pitch stimuli, in general, are processed by the pitch processor in the same way as assumed for monaural complex sounds (Bilsen, 1977). This central spectrum theory for binaural interaction has been further developed by Raatgever and Bilsen (1977) and Raatgever (1980). It describes phenomena like lateralization, dichotic pitch and binaural detection based on the generation of central activity patterns. In this, central spectra are supposed to be the results of an internal-delay dependent addition of the analyzed signals from both ears. The spectral dominance concept is included in the model by a dominant spectral weight of the 600 Hz region.

One of the most salient dichotic pitch phenomena is evoked by the MPS-noise stimulus (Raatgever and Bilsen, 1977; Raatgever, 1980; see also next section). According to the central spectrum theory, this stimulus gives rise to a central spectrum with sharp peaks at harmonically related frequencies. Raatgever (1983) investigated for MPS the position of the dominant region by comparing pitch matchings of this stimulus with those of the same but anharmonic stimulus. It was concluded that for MPS also a fixed frequency region around about 600 Hz is dominant. The aim of the present experiments is to compare the position of the dominant region for this dichotic noise stimulus with that for monaural noise stimuli with comparable comb spectra as well as with pure tones with cosine-shaped spectra in one set of experiments using the same experimental procedure and observers.

SIGNAL DESCRIPTION AND EXPERIMENTAL PROCEDURE

Three types of noise stimuli have been applied. Two stimuli have monaural complex spectra (Cosine-noise and Comb-noise) while one is dichotic in the sense that particular interaural phase shifts are applied (MPS-noise). From all configurations harmonics as well as anharmonic versions have been used.

Cosine-noise is a monaural (or diotic) stimulus that is made by delaying white gaussian noise over a time t and by adding, or subtracting, the undelayed noise. The power spectrum is then given by:

$$P(f) = 2 \times 2 \cos 2\pi ft$$

It is cosine-shaped with harmonically related peaks at n/t or with anharmonically related peaks at (2nt)/2t.

Like Cosine-noise, Comb-noise is a monaural (or diotic) stimulus. It is made by passing white noise through a delay line. A fraction g of the output is added to, or subtracted from, the input (feed-back). This leads to the power spectrum:

$$P(f) = 1 - g^2 (1 + \cos 2\pi ft)$$

The harmonic and the anharmonic spectra have sharp peaks at the same frequencies as for Cosine-noise.

MPS-noise is a dichotic stimulus with a flat power spectrum at both ears. The signal at one ear is the result of noise passing a delay line (delay t) with feed back of a signal fraction g. A fraction -g(1-g^2) of the input signal is also added to the output of the delay line. The signal at the contralateral ear is the input signal multiplied by (1-g^2). Do we suppose the binaural interaction process to be an addition, then the central spectrum is characterized by:

$$P(f) = \frac{2 \times 2 \cos 2\pi ft}{(1-g^2)}$$

It can be proved that the power spectra of Comb-noise and MPS-noise have the same shape apart from the spectrum of Comb-noise being a factor g less modulated and lifted over 1/(1-g^2).

Pitch matching experiments have been carried out using the stimuli described above. These stimuli were presented either diotically or dichotically at a level of 40 dB SL. Three observers took part in the experiments. The procedure was controlled by a computer. Signals, characterized by different t-values
(or peak spacings in the corresponding spectra), were presented in a random sequence with \( \tau \) varying between 2 and 10 ms. Subjects had to match the perceived pitch with the pitch of an adjustable reference stimulus. It consisted of a periodic pulse with a noise background in such a way that it had more or less the same character. The adjusted pulse frequency or pulse interval \( \tau \) was read by the computer. If the perceived pitch was ambiguous (anharmonic situations) the observers were asked to match the most salient or readily adjustable pitch.

**DISCUSSION AND CONCLUSIONS**

Pitch matchings using diotic or dichotic noise stimuli with harmonic central spectra have been performed just for completeness. Besides, they enable direct comparison with the same anharmonic stimuli. The results for diotic Cosine-noise and Comb-noise as well as for dichotic MPS-noise agree with the expected relation \( \tau_m = \tau \).

1. Is the dominant frequency region a fixed region or has it a relative position such that a particular harmonic number gives a better description?
2. If the dominant frequency is fixed: at which frequency, or if it is a particular harmonic (or harmonics): at which harmonic number is it located?

Since we have measured for different \( \tau \) of anharmonic noise the corresponding \( \tau \) of a harmonic stimulus that has the same pitch (e.g. Figs. 1, 2) we formulate the above questions in terms of relations between \( \tau \) and \( \tau \). For a fixed dominant region around \( \tau \), we have \( \tau = \tau \pm 1/2\tau \). The matchings are shifted over half the period of the dominant frequency resulting in 2 lines that are parallel with the lines representing the matchings of the harmonic stimulus. If the dominant region is described by a harmonic number \( n \) we expect: \( \tau = (1 \pm 1/2n)\tau \). The matchings can be described then by lines through the origin.

In Figs. 1, 2 for one subject the anharmonic matchings for 2 noise stimuli are given. It can be seen that the matchings for the dichotic noise (MPS-noise, Fig. b) are situated on lines parallel to the line for the harmonic case while for the monaural noise (Cosine-noise) the lines tend to pass the origin. This trend was found for all observers.

Tests using the method of least squares and comparing the different residual variances resulted in the conclusion that for dichotic noise signals with a comb-shaped central spectrum the fixed frequency region around 600 Hz on the average is dominant, while for monaural Comb-noise as well as Cosine-noise the 4th harmonic is dominant with a tendency to higher harmonics for some subjects.

Further, it has been found to depend on observer and stimulus type whether sometimes the pitch is dominated by the spectral pitch of the lowest component.

**REFERENCES**


---

From the pitch matchings of the above stimuli with spectra that are anharmonic in the way described before one should be able to draw conclusions on the position of the dominant frequency region.

We will limit ourselves here to the following questions for all stimuli used:
AUDITORY AFTER-IMAGES
PRODUCED BY COMPLEX TONES WITH A SPECTRAL GAP

Hugo Fastl
Institute of Electroacoustics, Technical University München, Arcistr. 21, D-8000 München 2, F.R. Germany

INTRODUCTION

About 20 years ago, Zwicker (1964) discovered the phenomenon that after switching off a band-stop noise, a faint tone may be heard. The effect was confirmed for band-rejected noise and pulse trains by Neelen (1967) as well as Luninis and Guttman (1972), who proposed to name the phenomenon a "Zwicker-tone". Despite the fact that a Zwicker-tone is the negative after-image elicited by a notched noise, its pitch strength or tonal quality is virtually identical to the pitch strength of a pure tone (Fastl and Stoll, 1979).

Since only relatively few data on Zwicker-tones have been published, we explored with 52 subjects the existence of a Zwicker-tone for a band-stop noise at 4 kHz, presented at various levels. The main issue of this paper, however, is to show that complex tones with a spectral gap also may elicit a Zwicker-tone. The dependence of the Zwicker-tone's pitch and loudness on relevant stimulus parameters will be discussed.

METHOD

All subjects participating in the experiments had normal hearing; their age was between 20 and 65 years. Only one subject had experienced the phenomenon of the Zwicker-tone before the start of the measurements. Sounds were presented monaurally in a sound isolated chamber via an electrodynamic earphone (Beyer D2 40) with a freefield equalizer (Zwicker and Flott scpeller, 1967). At the subject's disposal was a switch with three positions: in position 1 the band-stop noise or complex tone with spectral gap was presented, in position 2 this sound was switched off, and in position 3 a pure tone could be heard. The level of that tone could be controlled by an attenuator and the frequency of the tone was controlled via a helix-potentiometer by which the voltage to a VCO was varied, yielding frequencies between 200 Hz and 8 kHz. The level and frequency of the pure tone heard by the subject in position 3 of the switch could be measured by the experimenter outside the booth. Bandstop noise was produced by a filter cutting out of white noise a notch centered at 4 kHz with 1150 Hz bandwidth and steep filter slopes of 0.4 db/Hz. Complex tones with a spectral gap and fundamental frequencies of \( f_n \) 20 Hz or 200 Hz were used. They were generated by adding up pure tones of equal amplitude but random phase with frequencies of \( n \cdot f_n \) (n=1,2,3,...) up to a maximal frequency of 10 kHz. In order to produce spectral gaps in the complex tones, some pure tones within the sequence \( n \cdot f_n \) were omitted. The sum of the added pure tones was stored in a memory of a desk-top computer. For the experiments, the content of the memory was cyclically read out and D/A converted by a 12 bit converter.

EXPERIMENTS

Survey

The first experiment was intended as a survey with respect to the question: "How common is the Zwicker-tone phenomenon?" 52 naive subjects were presented white noise with a notch of 1150 Hz at 4 kHz, a bandwidth of 20 kHz and 43 dB overall SLP. The notched noise could be heard in position 1 of the switch and was switched off in position 2. The subjects were not informed about the signals presented in different switch positions, but were asked: "What do you hear, when switching from position 1 to position 2 of the switch?" 26 subjects (50%) answered spontaneously that they hear a faint, decaying pure tone. 10 subjects (19%) reported a Zwicker-tone after being informed "you may hear something which is soft" and 13 subjects (25%) reported a Zwicker-tone after the hint "you may hear something which is short". The remaining three subjects (6%), however, could not hear a Zwicker-tone even after an explanation of the stimuli and the expected phenomenon. Summarizing the results of our survey it can be stated that 94% of our naive listeners could hear a Zwicker-tone without information about the phenomenon.

With the same 52 subjects we explored, at which levels of the bandstop noise at 4 kHz, a Zwicker-tone can be produced. The overall SLP of the bandstop noise was varied in 5 dB steps between 23 dB and 83 dB and the question was: "Can you hear a Zwicker-tone?" We got 75% positive answers for levels between 38 dB and 53 dB, and 25% positive answers for levels between 23 dB and 73 dB. This is in line with data from the literature (Zwicker, 1964; Luninis and Guttman, 1972) that we found that in a large number of subjects bandstop noises with levels around 43 dB overall level (about 8 dB spectrum level) produce a Zwicker-tone.

Complex tones with spectral gap

It is known from the literature (Zwicker, 1964; Neelen, 1967; Luninis and Guttman, 1972) that line spectra produced by pulse trains, from which several lines are removed by filtering or cancelling, may elicit a Zwicker-tone. While the temporal envelope of these line spectra shows high peaks, we synthesized complex tones with a spectral gap, showing much less peaked temporal envelopes because of random phase relations between the sinusoidal components. With 9 subjects we measured the pitch and loudness of Zwicker-tones elicited by complex tones with fundamental frequencies of 20 Hz and 200 Hz, and a spectral gap of 1200 Hz at 4 kHz for 40 dB overall SLP.

Fig. 1: Zwicker-tones elicited by bandstop noise and complex tones with a spectral gap. Frequency f and sensation level SL of pure tones which produce the same pitch and loudness as the corresponding Zwicker-tones. (a) and (d) Bandstop noise with 43 dB overall level and a gap of 1150 Hz at 4 kHz. (b) and (e) Complex tone with 200 Hz fundamental frequency and 40 dB SPL; gap at 4 kHz is 1200 Hz wide. (c) and (f) Complex tone with 20 Hz fundamental frequency, 40 dB SPL and 1200 Hz-gap at 4 kHz.
The subjects first listened to the complex tone with spectral gap, switched it off to hear the Zwicker-tone, and then adjusted a pure tone in frequency and level in such a way that the pure tone produced the same pitch and loudness as the Zwicker-tone. Results of this experiment are given in Fig.1 together with data for bandstop noise. The left panels show the frequency f and the right panels the sensation level S1 of the pure tone matched to the respective Zwicker-tone. The symbols represent individual medians, each calculated from eight adjustments which coincide within about 2% on the average; the numbers with bars indicate the corresponding global medians with interquartiles. The data depicted in Fig. 1 show quite similar results for bandstop noise (a) and complex tones with a fundamental frequency of 200 Hz (b) or 20 Hz (c). Despite individual differences, results plotted in the left panels of Fig. 1 indicate that, on the average, the pitch of the Zwicker-tone corresponds closely to the centerfrequency of the spectral gap. Data plotted in the right panels of Fig. 1 also show pronounced individual differences in the sensation level of pure tones producing the same loudness as Zwicker-tones. On the average, for bandstop noise, a sensation level of about 10 dB is reached (d) while for complex tones with spectral gap (e) and (f), a sensation level of about 12 dB shows up. This larger sensation level corresponds to a larger loudness of the respective Zwicker-tones and would seem to confirm the reports of the subjects that the Zwicker-tone is more distinct and stable for complex tones with a spectral gap than for bandstop noise.

Since the existence of Zwicker-tones for complex tones with a spectral gap was established, we explored with 5 subjects the minimal width of the gap necessary to produce a Zwicker-tone. At 2 kHz centerfrequency, we found 400 Hz minimal gap width for complex tones with 200 Hz fundamental frequency and a 240 Hz gap for 20 Hz fundamental frequency. At 4 kHz centerfrequency, the minimal gapwidth is 800 Hz for 200 Hz fundamental frequency and 400 Hz for 20 Hz fundamental frequency. Thus, the minimal gapwidth necessary to produce a Zwicker-tone with complex tones depends both on the centerfrequency of the gap and the fundamental frequency of the complex tone, i.e., the spacing between the spectral lines. For each complex tone with line spacings of 200 Hz, the spectral gap has to exceed one critical band to elicit a Zwicker-tone, while for complex tones with 20 Hz line spacing spectral gaps only half a critical band wide can produce a Zwicker-tone.

In the last experiment we discuss here, we used complex tones with 20 Hz fundamental frequency, 10 kHz bandwidth and 40 dB overall SPL and introduced spectral gaps of different widths. Fig. 2 shows results for gaps around 4 kHz in the upper part and for gaps around 2 kHz in the lower part. The symbols indicate individual medians of four subjects, each derived from four pitch matchings. Again, the reproducibility of the pitch matches was good, the average deviations amount to only 1%. From the data displayed in Fig. 2 it becomes clear that for small spectral gaps, the pitch of the Zwicker-tone corresponds to the centerfrequency of the gap, while for larger gaps the pitch of the Zwicker-tone drifts away from the center towards the low-frequency edge of the gap. For a constant width of the gap, this effect depends on the centerfrequency of the spectral gap: For an 800 Hz gap at 4 kHz (first line in Fig. 2) the pitch of the Zwicker-tone corresponds to the center, whereas for an 800 Hz gap at 2 kHz (sixth line in Fig. 2) the Zwicker-tone occurs at lower frequencies closer to the low-frequency edge of the gap.

**DISCUSSION**

In this paper we showed that complex tones with a spectral gap may elicit an auditory after-effect, called Zwicker-tone. Much more work is necessary to explore the behavior of Zwicker-tones in detail. However, it seems already to be justified to distinguish the Zwicker-tone phenomenon from other after-effects of sound coloration (see e.g. Wilson, 1970; Vieille, 1970; Terhardt, 1980). The result that with increasing width of the spectral gap, the Zwicker-tones are shifted from the center towards the low-frequency edge of the gap would seem to suggest an edge effect for the description of the phenomenon (see Lumis and Guttman, 1972). However, it seems unlikely that an edge effect is observed for the pitch of low-pass noise (Fastl, 1978) may fully account for the Zwicker-tone phenomenon, since we know (Fastl and Stoll, 1979) that the pitch strength of low-pass noise is by a factor of five smaller than the pitch strength of a Zwicker-tone.

**References**


Fastl, H. and Stoll, G. (1979), Hearing Research 1, 293-301.


Zwicker, E. and Feldtkeller, R. (1967), Das Ohr als Nachrichtenempfänger, Hirzel-Verlag, Stuttgart.
ON THE DEFINITION OF TONE COLOR

O. Kitamura

Osaka University of Arts, 3, Nishishinmachi, Kannoncho,
Minami-ku, Osaka, 555 Japan

It is called that loudness, pitch and timbre are
the three attributes of auditory sensation. American
National Standards (1970) defined timbre as following.
"Timbre is that attribute of auditory sensation
in terms of which a subject can judge that two sounds
similarly presented and having the same loudness and
pitch are dissimilar." Generally, this definition is
typical one of timbre at present, and it is supposed
that a part of description on "Klangfarbe" by Helmholtz
was picked up and it was followed in the footsteps.
Although in many cases of research on tone color, we
encounter the case which we can't carry out the re-
search if ANS definition of timbre is obeyed strictly.
For example, when the pitch of tone of a musical
instrument is changed, we perceive not only the pitch
difference between the former and the latter, but the
difference of tone color between both. In this case,
if we obey the above mentioned definition of timbre
strictly, we must not compare these two tones with
regard to the aspect of tone color decided by the
definition. Moreover, as Risset et al. mentioned,
when the intensity, the frequency and the frequency
spectrum of a saxophone were changed to a considera-
ble degree, we can identify the saxophone.
On the other hand, there is an opinion that
timbre is the whole of tone quality other than loud-
ness and pitch. But this is ambiguous opinion.
Under those circumstances, several researchers of
tone color entertain a doubt about ANS definition of
timbre (1). (6)

One of the objects of research on audible sound
is to make clear the relation between hearing and
physical properties of sound, the unit of loudness
is tone and the unit of pitch is me but there was
no unit of tone color which can be used for quanti-
tative measurement of tone color. Many researchers
have researched to find the scales which can make
quantitative measurement of tone color possibly.
More than 50 studies were carried out to re-
search the number of factors of tone color about all
kinds of sound by the author and other people; with
the purposes (1) to make clear the meaning of tone color
and (2) to make the descriptive adjective scales
which can measure tone color quantitatively.
The methodology employed utilized the semantic differen-
tial method developed by Osgood. As the results of
these research, three principal orthogonal factors
of tone color were extracted. Namely, factor of beautifullness to which beautiful, ugly, clean, etc.
belong; factor of powerfullness to which powerfull,
feeble, etc. belong; factor of metallic impression to
which deep, metallic, etc. belong. In several experiments, one more principal factor of tone color
was extracted.

The impression of sound expressible by these
descriptive adjectives of tone color should like to
be named the sound impression about descriptive ad-
jective of tone color. This sound impression about
descriptive adjective of tone color is the different
aspect of sound impression from the aspect which can
identify the sound source, for example, sound of
violin, sound of piano, sound of flute, sound of car,
sound of thunder, etc. The sound impression about
descriptive adjective of tone color is changed when
the physical properties of sound (for example, frequency
spectrum, growth, etc.) change, but the aspect of
identification of sound source is not changed in many
cases.

According to these consideration on tone color,
it is concluded that it is reasonable to replace the
auditory sensation called timbre or tone color by the
following two aspects of sound.
(1) aspect of identification of sound source,
(2) aspect of sound impression about descriptive
adjective of tone color.

As mentioned above, Helmholtz explained the tone
color by means of identification of musical instru-
ment. This explanation is about identification of sound
source. In another place of "The sensation of tone"
Helmholtz explained that, using same musical
instrument, it is possible to play several kinds of
tone color and that the sound impression is changed
as soft, poor, rich, dull, bright etc. according the
changed playing. This is the explanation concerning
sound impression about descriptive adjective of tone
color.

It is desired to renew the concept and defini-
tion of timbre or tone color and decide the definition
of tone color according above mentioned ideas and
independent to loudness and pitch.

(1) Helmholtz, H.L.F. von "Die Lehre von den
Tonempfindungen alla physiologische Grundlage für
die Theorie der Musik", F. Vieweg & Sohn,
Braunschweig. erste ausgabe (1913), p19-20

(2) Risset, J.C. & Wessel, D.L. Exploration of
Timbre by Analysis and Synthesis [D. Deutsch;
The Psychology of Music] p26

(3) Sakai, T., Chō-kaku to Onkyō-shinri, Corona
Press (1970), p74

(4) Plomp, R. Timbre as a Multidimensional Attribute
of Complex Tones, [Plomp et al; frequency analysis
and periodicity detection in hearing] A.W.
STIJNHOFF LEIDEN (1970)

(5) Plomp, R. Aspects of Tone Sensation, Academic

(6) Kitamura, O., Namba, S., and Matsumoto, R. (1968)
Factor analytical research of tone colour, Rep.
6th Int'l Cong. Acoustics, Tokyo, Vol.I, ppA17-
A120. Elsevier, Amsterdam.

(7) Helmholtz, H.L.F. von "Die Lehre von den
Tonempfindungen alla physiologische Grundlage für
die Theorie der Musik", F. Vieweg & Sohn,
Braunschweig, sechste ausgabe (1913), p31

(8) The same as the above, p118-120
SPATIALLY DEPENDENT FREE FIELD MASKED THRESHOLDS

R.W. Gatehouse
Psychology Dept., University of Guelph, Guelph, Ont.
Canada N1G 2W1

INTRODUCTION

Masking, failure to detect or reduced detection of a signal(s) in the presence of an or other sounds has been widely investigated and its many facets have been the subject of numerous reviews (e.g., Yost & Nielsen, 1977; Carterette & Friedman, 1978). Since these aspects are well known it is unnecessary to delineate and reference them here. Rather, it is sufficient to indicate that masking degree variants have been shown for differing signal types, temporal factors, and presentation methods such as binaural, monaural, continuous or non-continuous et cetera.

Despite the wealth of information gathered via such procedural variants, it can be argued that two procedural variations have been largely overlooked. First, nearly all data have been gathered under headphone rather than free-field presentations of signals and maskers. And, second, that relatively little is known about the effects produced when signal(s) and masker(s) are spatially separated. It can of course be also argued that such conditions have been adequately mirrored in BMLD studies wherein for example, a binaural masker can be presented against either a monaural or a signal out of phase. In essence, conditions wherein no interaural differences or free-field 0° presentations, are posed against signals with spatial differences. That such studies give evidence for psychophysical tuning curves and all that such analogues of neural tuning curves mean (i.e. tending to particular CF’s; steep high frequency cut-offs; etc., e.g., Green & Yost, 1975; Green, Shelton, Picardi & Hafter, 1981) is also well known. Nonetheless, free-field studies in which the masker(s) and signals are spatially separated as in the everyday listening environment are needed at minimum to confirm, support, or negate such earphone data. Two studies which have separated the maskers and signals are those of Cavan et al. (1979) and Kurozumi and Ohashi (1981). In the former, subjects in an anechoic room listened to a masker (either 55 or 45 dB SPL) from 0°, and 20 msec. later to a 1.0 kHz signal from a left (-15°) speaker. The sequence was then repeated with the signal from a right (+15°) speaker. The masker frequency was varied (0.1; 0.5; 0.7; 1.0 or 1.3 kHz) and the task was to judge whether the sound ensemble percept changed with the signal position change. The authors reported that signals are more difficult to localize in the presence of maskers which are in the same critical band. In the latter study, 62dB SL WN maskers (50 msec) were presented at 0, 30, 90, 120 or 180 degrees relative to shorter WN signals from similar positions to the left or right of 0° on single channel speakers. As well, two channel masking from +90° was tested. Threshold measurements were obtained. The results showed that the degree of masking was highly dependent upon the position of the masker. That is, masking was greatest when the masker and mask she coincident, decreased for azimuths out to 45°. Beyond 90° separations of the masker and signal, the functions were asymmetrical with some loci showing different amounts of masking (up to approximately 10 dB less attenuation) than others. The functions however were symmetrical about 0° and 180°. In the latter case (masker at 180°), mask levels varied only by 4-6 dB for signals at positions between 190°. But, for positions between +90° and 180°, there was greater masking of the signals than when they were coincident at 180°.

The present study determined the masking of tonal signals at loci from 0° to +90°, by a WN masker fixed at 0° azimuth.

METHOD

Subjects: Four normally hearing 25 year old subjects. All were given binaural a.c. testing .25 to 8.0 kHz; av. loss in each ear < 15 dB.

Apparatus and Procedure: Ss were seated in a height-adjustable pedestal chair that permitted their ears to be aligned with axial plane speakers which were 1.5 m above the floor of the room. Heavy curtains, 182.9 cm, in height were mounted on a frame (3 vertical slats at 13° and 22° respectively, joined at the top and at floor level) of 12.8 mm diameter polyvinyl tubing. Subjects were thus visually isolated from the sound delivery system and room. Curtains also covered the top of this area. After task instructions subjects were fitted with a headband mounted light, and shown how to signal their responses via a two-button hand-held control box. The light, if unaligned with a photo-cell embedded in the 0° strut, interrupted the signal programme, and thus eliminated head movements. The curtained space was lighted by a 25 watt bulb. External to this enclosure directly above the subject’s head, was a computer controlled, ceiling mounted, motorized boom trap system whereby a single TDH-39 speaker could be positioned within (15°) of selected (azimuths, 1.1 m from S’s head position. A standmounted speaker was positioned at the same distance from S at 0° so that when it was coincident with the boom-mast approximately 1 cm separated them. All of this apparatus was situated in a sound treated room.

The signals (WN masker passed < 20 kHz., and various tonal signals) were generated by computer-controlled standard acoustic equipment housed in an adjacent room. Outputs from this latter equipment were controlled via a programmable attenuator.

The study proceeded in two steps. First unmasked threshold measures were obtained for the tonal signals (0.25, 0.5, 1.0, 2.5 and 4.0 kHz) at each of 0°, 30°, 90°, +30°, +45° and 90° azimuths. Next, 50 msec tonal probes of 90 dB SPL at speaker position, were delivered centered within one of two 100 msec bursts of the WN masker also delivered at 90 dB SPL from the 0° position. The WN bursts were separated by a 300 msec. silent gap. Subjects indicated by depressing the left or right button of the hand-held control, which WN burst he/she thought the tone was buried in. A 2 AFC paradigm in which two successive correct discriminations initiated a 5 dB increase in signal level until 4° turnarounds were noted or a 2 dB increase to an incorrect response. Thereafter, decreases too were 2 dB. Ten threshold runs were performed for each frequency at each position, and the average of these 10 was considered spatially dependent threshold (see Levitt, 1971; Green et al., 1981). Since the procedure took considerable time (~ 1 1/2 hrs. per session) testing was done over several days.

RESULTS

Averaged mean thresholds across subjects (masked vs unmasked for 5 frequencies by positions) were analyzed by separate ANOVAs. Paper-length limitation preclude separate presentations of these, but all frequencies masked vs unmasked and positions were significant factors (p < .05) and two-way interactions between these variables were obtained. Instead, the averaged masked threshold functions alone are presented (Fig. 1). The ordinate shows the average attenuation level decreasing 90 dB SPL at which the signal could be detected for the various azimuths (absissa). For clarity no a.d.’s are presented but
individual thresholds ranged over 9 dB for some positions (average 5 dB). As can be seen no consistent relationship between best-detected (lowest attained threshold) position and masker frequency exists. For example, a masked 1.0 kHz signal, which in general has lowest masked thresholds, is best detected at positions where signal and masker are coincident (0°) and, farthest apart (+90°) with decreasing detectability from 0° to +30° followed by a "turn around" and increasing detection to the periphery. A similar pattern is shown for 2.0 kHz but with the poorest threshold and "turn around" at +45°. The other tonal signals in general are more difficult to detect at or near 0° where signal and masker are coincident, with increasing detectability as the signal and masker diverge. The exception is at 250 Hz with "turn arounds" occurring at two positions (+30°; 45°); especially evident for loci to the left of 0° but on the whole somewhat improved detection as the signal diverges from the masker. Finally, it might appear that there is considerable difference in attenuation levels between best- and least-detectable positions across frequencies, but this is not so. The approximate differences were: 0.25 kHz = 13; 0.5 kHz = 14.5; 1.0 kHz = 12.5; 2.5 kHz = 15.5 and 4.0 kHz = 14.5 dB.

**DISCUSSION**

The masked thresholds are notable in several ways. It was expected that as the signal binaural differences were emphasized (i.e. increased angular displacements from coincidence with the fixed (0°) masker) that detection or localization would similarly be increased. This occurred only for 0.5 and 4.0 kHz signals, partially for 0.25, and not at all for 1.0 and 2.5 kHz. For the latter two, detection was nearly equivalent at masker-signal coincidence as at greatest interaural difference (+90°). These observations are not entirely consistent with dichotically derived data i.e. increasing masking release and improved detection (Hirsch, 1948; Green & Henning, 1969; Hafer & Carrier, 1970). The masked function shapes are also notable. For 0.5 and 4.0 kHz, increased detectability to +90° are akin to Kurozumi & Ohgushi's (1981) WN/WM functions. For 1.0 and 2.5 kHz signals the symmetrical "W"-shaped functions, mirror those of the same authors, except the "turn arounds" occur within, rather than outside of the +90° positions. The reasons for discrepancies between these and other data are not readily apparent.

The 4.0, 0.5 and in part 0.25 kHz data suggest that free-field detection and localization is not fundamentally different from earphone obtained thresholds —increasing interaural differences are accompanied by increased detectability even in presence of a fixed masker at 0°. Explanation difficulties are mostly seen (1.0 and 2.5 kHz) where thresholds increase to +30° or +45°, and this suggests a new type of increased binaural cues. Perhaps, the head-movement control (head-light assembly) was not effective that it eliminated them, but this would overlook other cues (pinna; hair; etc.), and why then, was it frequency selective. "Cone of confusion" effects beyond +45° (Mills, 1972; Searle et. al., 1976) are also counter indicated since increased detection was observed here. Therefore, increased/decreased detection to/from turn-arounds must be due to pinna transforms, or selective attention (e.g. cocktail party) effects where increased detection does not necessarily follow from signal and masker separations. Finally, it would be interesting if all of the tonal signals had similar "W"-shaped masking functions (as 1.0 and 2.5 kHz), ordered themselves in overall detectability by frequency, and by sharpness/flatness. Then their consideration as spatial analogues of psychophysical tuning curves could have been continued.

It can be concluded that free-field detection and localization under masking conditions, may be different from earphone-derived analogues. The former seem more subject to frequency and positional variants. A similar suggestion has been made by others (e.g. Carhart et. al., 1969). Further investigation is underway to see what occurs if: the pinnae are occluded; testing is done beyond +90°; signals instead of maskers are fixed; and, impaired subjects (e.g. monaurals) are tested.

**REFERENCES**


LATERALIZATION SHIFTS AS A FUNCTION OF INTERAURAL DELAY AND TONE DURATION

C. Baruch, S. Giani, M. -C. Botte

Laboratoire de Psychologie Expérimentale, CNRS, 28 rue Serpente, 75006 Paris, France

INTRODUCTION

The classical data (Yost, 1961; Blauert, 1983) on lateralization has shown that for two dichotically presented pure tones of frequency less than 1500 Hz, a phase shift associated or not with an interaural delay produces a lateralization of the sound image towards the ear leading in phase. This lateralization is greatest for frequencies below 800 Hz (1/2 period \(= 643 \mu s\)).

A recent study by Rusani and Poli (1985) compared the just noticeable difference in lateralization with three types of presentation: a simple phase shift with no interaural delay, an association of interaural phase shift and delay, or interaural delay with no phase shift. In the first two cases their results agree with the previous data. However, a very high jnd was observed in the third condition, apparently indicating that the auditory system is relatively insensitive to interaural delay when it is not accompanied by a phase shift.

We have previously shown (Botte, Baruch, and Scharf, in press) that with long interaural delays the lateralization of two tones with the same frequency and presented in phase is somewhat unexpected. On the beginning of the second tone, 20 seconds after the onset of the first, a single fused image was heard but this image was not heard in the center of the head; it was lateralized, for a few seconds at least, towards the ear where the second tone was presented. This decentralization, as opposed to the centralization with no interaural delay, occurs for tones with an interaural delay difference but was more distinct with tones of similar frequency. We called this phenomenon "paradoxical lateralization" because the auditory image was lateralized towards the ear that was stimulated second.

The present experiment was designed to specify the conditions in which this paradoxical lateralization appears. Firstly, we tried to find if there was a critical interaural delay from which the image of tones presented in phase would move from their central position, observed for delays of less than 1 ms, towards the ear stimulated second, as already observed for 20 ms interaural delay. Secondly we wished to examine the influence of the duration of the two tones' coexistence to see if this factor would eventually interact with the length of the interaural delay.

PROCEDURE

Tones were digitally generated in real time by a synthesizer (DMX 1000) piloted by a microprocessor (Northstar Horizon). Levels were controlled by decade attenuators (laboratory prototypes). Sounds were presented to the subjects in a sound-proof room, through earphones (TDH 49).

Two 1000 –Hz, 60 –dBA tones were presented, one to the right ear, the other to the left. Right ear stimulation began after left ear stimulation with a variable onset delay (0, 0.5, 1, 2, 5, 7, 10, 25, 50, 100, 500, 1000, 5000, 10000, or 20000 ms); and the duration of the two tones' coexistence was either 40, 200, 500, 1000, or 2000 ms. The two tones always stopped together and their cosine-squared slopes lasted 10 ms.

In one block of trials the 15 interaural delay conditions were presented in random order for the same coexistence duration. Each experimental session consisted of 12 blocks of trials with the same coexistence duration. Each subject received 5 experimental sessions corresponding to the 5 different coexistence durations.

The subjects had to estimate the intracranial position of the sound image at the beginning of the second tone, and represent it by means of a movable arrow on a linear potentiometer. This arrow could be turned a total of 180° in the form of a semi-circle. After each trial the subject had to reset the arrow to the median position. The estimation was transformed into degrees: 0° representing a central position in the head, 90° a lateralization to the extreme right, and –90° a lateralization to the extreme left.

On each trial (for a given coexistence duration and interaural delay), the stimulus was presented 5 times in succession and the subject's estimate of lateralization was recorded at the end of the 5 presentations. Altogether, we therefore obtained 12 estimates per subject for each of the 75 stimuli used.

Four audiometrically normal subjects between 23 and 27 years of age took part in the experiment after undergoing a long training session.

RESULTS

The results averaged over the 4 subjects are given in Figure 1.

![Figure 1: Lateralization of sound image after the beginning of the second tone for 1000 Hz, 60 dB SPL, tones as a function of interaural delay. Each curve represents one of 5 different coexistence durations for the two tones that terminated simultaneously. The left ear was stimulated first; 90° represents an extreme right lateralization, –90° an extreme left lateralization, and 0° a central image. Mean of 4 subjects.](image-url)

The lateralization estimates are represented as a function of interaural delay for each of the 5 coexistence durations tested. Although the individual subjects scores differed quite a lot, and
particularly for interaural delays between 2 and 25 ms, the four subjects gave the same general form of curve.

The following points were observed:

- An interaural delay less than or equal to 1 ms does not modify the position of the auditory image compared to a zero delay. The image remains centered.

- There exists a critical interaural delay above which the image shifts towards the ear stimulated second. Starting from a delay of approximately 25 ms, a lateralization is observed that becomes greater as the interaural delay increases. For three of the subjects, above a 25 ms delay, the lateralization is significantly shifted towards the right side that is towards the lagging ear; for the last subject, the decentralization is significant only for the delays of at least 50 ms.

- Individual differences are large between 2 and 25 ms delays. Two of the subjects showed a significant lateralization towards the ear stimulated first, while the other two subjects continued to estimate the image's position as central. These results may be explained by the existence of a temporary difference in intensity of the two tones due to their rise time (10 ms). On the onset of the second tone, a lateralization towards the ear stimulated with the greatest intensity, that is the ear stimulated first, might therefore be produced.

- The intra-individual standard deviations are considerably higher (as much as 20°) with delays between 2 and 10 ms. This may be explained by the fact that at these delays the image becomes more diffuse and so its position is more difficult to estimate.

- There is no systematic interaction between interaural delay and coexistence duration as shown by an analysis of variance. Whatever the coexistence duration the critical delay remains the same.

- For the delays lasting more than 25 ms, lateralization is significantly greater when the two tones coexist for 500 ms or less. This may indicate that lateralization towards the ear stimulated second is a relatively transitory perceptual phenomenon. When the coexistence duration was sufficiently long, a decentralization was presumably occurring as observed in previous experiments with coexistence duration of 10 s. If the coexistence duration is 2000 ms, as in the present experiment, time is allowed for some amount of decentralization; since the estimate of lateralization is asked for only at the end of the coexistence period, the decentralization is apparently lesser in that condition.

CONCLUSIONS

The present results confirm the existence of "paradoxical lateralization" which is produced even when the two tones coexist very briefly.

This lateralization towards the ear stimulated second appears with an average critical interaural delay of 25 ms, and when this delay is increased, that is with longer presentation times of the first tone alone, the lateralization (or decentralization) caused by the onset of the second tone also increases.

It appears that 25 ms might be the time required by the auditory system to begin processing the second tone as "new" compared to the first tone, even if the two tones have the same frequency, the same intensity, and are in phase. These results may be related to the data on contralaterally induced adaptation (Botte et al., in press) that show a large reduction in loudness of a 1000 Hz tone, already present for 70 seconds in one ear, when a second tone of the same frequency and level is presented in the other ear. The lateralization phenomenon that we have demonstrated here seems to be linked to this reduction in loudness. Its functional role may be to make each new sound stand out from an already existing background sound.

REFERENCES


THE PRECEDENCE EFFECT REVISITED: ECHO SUPPRESSION WITHIN AND ACROSS FREQUENCY BANDS

P. L. Divenyi and J. Blauert

Speech and Hearing Research Facility, Veterans Administration Medical Center, Martinez, California, 94553, U.S.A., and Lehrstuhl für Allgemeine Elektrotechnik und Akustik, Ruhr-Universität, Bochum 1, Federal Republic of Germany

INTRODUCTION

The "precedence effect", "Haus effect", or "law of the first wavefront" (Wallach et al., 1949; Haus, 1951) refers to the phenomenon whereby a stimulus originating at one point in the auditory space has the capacity to suppress the perception of its echo with a different spatial locus. Recently, this phenomenon has received renewed attention (Zarek, 1980; Cowan, 1985; Blauert, 1984). These studies have shown that the precedence effect is a potent factor in binaural signal processing in that it can effectively disrupt the two major binaural processes underlying auditory localization: interaural time- and interaural intensity discrimination.

One aspect of echo suppression, however, has not been investigated. It is not clear to what extent echo suppression is restricted to the frequency band of the first sound or whether its effect transcends across bands. In other words, we should ask if a sound that was first localized will suppress the perception of only a delayed copy of itself or whether it will also block out the spatial percept of another sound arriving later from a different spatial position. The present study represents an effort to elucidate this problem.

METHODS

In all experiments reported below the stimulus consisted of a brief conditioner (CD) sound followed by a brief target or probe (PR) after a few milliseconds separation, delivered dichotically through earphones (Sennheiser HD240). The CD was always diotic whereas the PR was either diotic or it arrived in the left earphone after a variable interaural time delay. The three trained listeners were asked to discriminate this interaural time difference (ITD) in a multiple-level two-alternative forced-choice procedure. The goal of the experiments was to determine, for each observer in each condition, the ITD corresponding to a 75% correct performance level. All stimuli were generated digitally (DEC PDP 11/34 computer) and delivered in synchrony through a pair of 16-bit D/A converters (Charybdis Analog I/O System) at a 33.33 kHz sampling rate. One subject was tested at a time in a single-walled IAC sound-treated chamber. Each reported threshold estimate, whenever such a threshold could be determined, is based on at least 480 trials.

RESULTS

1. Experiments with Filtered Clicks

Stimuli for this series of experiments were 33-usec clicks either low-pass (f₁ = 1.5 kHz) or high-pass (f₁ = 2.5 kHz) filtered by means of 6th-order Butterworth filters. All clicks were normalized with respect to peak-to-peak amplitude. In the control experiment, ITD threshold for a high-pass click (PR in the absence of any CD) was first determined. In the next experiment, the same PR was preceded by a low-pass CD delivered either 4 or 6 ms prior to the PR (onset-to-onset); under these conditions the ITD threshold could not be measured, i.e., localization of the probe was suppressed when the PR preceded the CD. When, on the other hand, the CD was the high-pass and the PR the low-pass click (with a 4-ms echo delay), the probe was clearly localizable. Figure 1 illustrates the results of one representative subject.

2. Experiments with Filtered Noise Bursts

Filtered clicks may not be the most appropriate stimuli for studying dichotic echoes because the interaction of the ringing of the CD and the first portion of the PR could produce unwanted temporal cues. For this reason, we opted for investigating the phenomenon more fully with filtered and undelayed bursts of white noise resulting from a CD - PR interaction would be thus minimized as long as the two bursts were independent noise samples. The first experiment in this series replicated one key condition of Zurek's (1980) study and the PR was an independent 1-msec broad-band noise burst with a 3-ms echo delay. In agreement with Zurek, we could not find evidence for probe localization.

In the experiments that followed we generated stimuli by passing the noise through critical-band filters and shaping the output with a Gaussian time window. The rms power of all stimuli was held constant. Three center frequencies were tested: 500 Hz (with a 5 ms window length), 1.5 kHz (3 ms), and 4.5 kHz (2 ms). The PR was the CD. With 1.5-kHz bursts, there was some localization at long (6-ms) echo delays but the echo was suppressed at a short (3-ms) echo delay. When the PR was a 4.5-kHz burst, an interesting suppression effect could be observed: With a spectrally identical (4.5-kHz) CD, lateralization of the echo was completely suppressed, localization was diminished with a 1.5-kHz CD, but was present with a 0.5-kHz CD. Some lateralization could also be observed when the CD was a 1-ms unfiltered click. (See results of the same subject in Figure 2.) These data suggest that if the probe has spectral energy not contained in the conditioner, echo suppression will not occur, unless the energy of the CD is concentrated in a region of the spectrum just below that of the PR. This is to say that echo suppression spreads upward in the spectrum.

The effect of spectral commonality can best be studied with noise stimuli having periodic spectra (comb-filtered white noise) that have a type of dual structure that arises naturally in a reverberant environment (Zurek, 1980). In particular, we investigated one extreme case in which both the CD and the PR (independent) noise bursts had spectra with the same period (1.5 kHz), except that one of them was harmonic and the other inharmonic. Since spectral peaks in the CD correspond to notches in the PR and vice versa, we expected no echo suppression in this condition and our data confirmed this expectation. When, however, we high-pass filtered both the CD and the PR above 2.5 kHz, lateralization was obliterated. These results strongly suggest that a lack of spectral coherence in the CD and the PR allows the echo to be perceived aurally. Thus, the spectral singularities of each sound can be peripherally resolved. Further evidence for the importance of peripheral frequency analysis in echo suppression was offered by another experiment in which the CD was a brief (10-μsec) band-reject white noise burst with a steep (8th-order Butterworth) notch between 0.5 and 3.5 kHz. When the PR was a band-pass noise with energy between 2.5 and 3.5 kHz, clear lateralization was observed. When, on the other hand, the PR center frequency was shifted
to 1.3 kHz, lateralization became more difficult (Fig. 3).

DISCUSSION

Our results indicate that the precedence effect persists as long as there is a spectral overlap between the first sound and the echo. However, the data also make it clear that the spectral overlap is not to be interpreted in the physical sense but, rather, in terms of frequency selectivity in the peripheral auditory system. Incoherent spectral singularities in the conditioner and the echo will allow the echo to be detected only if these singularities can be resolved. Further, suppression of an echo appears to follow a spectral pattern similar to masking in that it spreads upward but not downward.

A similar dependence of binaural interaction on peripheral frequency analysis has been reported for binaural fusion (Scharf et al., 1976; Ebata and Sone, 1968) and for saturation effects in lateralization (Hafer and Busel, 1969). From the point of view of the auditory system, such a dependence makes profound sense: An echo should be suppressed only if in reality it is the echo of a sound first heard directly. The architectural acoustician will not be surprised to learn that a broad-band sound object is more effective in suppressing higher-frequency echoes which could interfere most with the perceived direct source of the sound (Blauert 1984). There could be a temptation to assign echo suppression to monaural forward masking. However, it must be stressed that, while the presence of the probe is clearly detectable, it is its spatial position that cannot be recognized. Thus, a masking interpretation of echo suppression should make it imperative to postulate a novel definition of peripheral masking, one in which not only the magnitude of sensation but also the resolution of the fine temporal structure of the waveform is considered.

ACKNOWLEDGMENTS

This research has been supported by the Veterans Administration, the Deutsche Forschungsgemeinschaft, and by a NATO Travel Fellowship.

REFERENCES


INFLUENCE OF AN INTERFERING NOISE ON THE LOCALIZATION OF A BAND NOISE SOURCE.

SUZUKI, Ybiti and SONE, Toshiro

Res. Inst. of Electr. Comm., Tohoku University
2-1-1 Katahira, Sendai, 980 JAPAN

1. INTRODUCTION

As to the effect of noise on sound localization for a tone, we reported in the previous ICA that the sound images shift to the direction opposite to that of noise when the signal frequency is about 1 kHz or below. As the consecutive study, the effect of noise on sound localization of a band noise source is studied.

2. EXPERIMENTAL PROCEDURE

Fig. 1 shows a schematic arrangement for the experiments in a anechoic room. Thirteen loudspeakers were placed within the angles from -60° (the minus sign means left to the front) to +60° (the plus sign means right to the front) every 5°.

An interfering noise was radiated by a loudspeaker whose direction was -30°, 0°, or +30°. The noise was comprised of a band-limited pink noise (300 - 8 kHz). The noise level was 50 or 70 dBSPL. A signal was either a 1 octave band noise (center frequency: 500, 1k or 2k Hz) or a band-limited pink noise ranging 300 - 8 kHz (i.e. the same as the interfering noise). The former was presented to create a real image or a virtual image, while the latter was presented to create only a real image. The real image signal was radiated from a loudspeaker located every 10° within ±30° with a sound pressure level of 50, 60, or 70 dB. The virtual image signals, on the other hand, were radiated in phase from a couple of loudspeakers located at -10° and +10°. One of the two channels is taken as the standard channel here. Futhermore, the sound level of the signal is indicated by the level of the standard channel (standard level) and the level of the other channel relative to the standard (relative level).

The standard level was 60 dBSPL and the relative level ranged from -15 dB to +15 dB.

Fig. 2 shows a schematic diagram of stimulus presentation. In the first stage, the signal sound is presented with the interfering noise. In the second stage, only the signal of the same sound level as the signal which appeared in the first stimulus was presented from a loudspeaker as a comparison stimulus. The subject could select one of the loudspeakers at his hand to adjust the direction of the comparison stimulus so as to coincide with that of the signal in the test stimulus. After that, the same series of stimuli were presented for the subject to judge once again. Then the result of his adjustment was read by a computer. The rise and decay times of the sounds were set to 100 ms so that clicks could not be heard. An stimulus condition was repeated five times. The subjects were four adults with normal hearing acuity.

3. EXPERIMENTAL RESULTS FOR A REAL SOUND IMAGE

Figs. 3 - 7 show the examples of the results. The abscissa shows the direction of the loudspeaker from which a signal sound is radiated, and the ordinate indicates the direction of the loudspeaker selected by the subject for a comparison stimulus. The broken line in the figure shows the judged direction in accordance with the direction of test signal. Marked values are the averages among the four subjects, since they showed similar tendencies.

Fig. 3 is an example to show that the subjects' judgements are very accurate in the noiseless condition. When the signal was the 1 octave band noise, it can be seen from Figs. 4 - 6 that a sound image for a test signal moves from its actual position to the opposite direction of the noise. The higher the noise level was, the greater the amount of the angle shift of the sound image was. The amount of the shift was greatest in 2 kHz and smallest in 1 kHz.

Fig. 7 shows the results for the band-limited pink noise. The sound image also moves from its actual position to the opposite direction of the noise when SN ratio is greater than or equal to 0dB. When SN ratio is -10 dB, however, the tendency of the results varies among subjects and they can be classified into two types. The first type of subjects tended to perceive that the signal sound moved away from the direction of the noise. This is the same tendency noted above. The second type showed a reversed tendency; the signal approached to the noise. A fusion of the images occurred in the case.
4. EXPERIMENTAL RESULTS FOR A VIRTUAL IMAGE

Figs. 8 and 9 show the examples of the experimental results. As all of the subjects exhibited the same tendency, the averages are shown in the figures. In these figures, the abscissa denotes the relative level. The ordinate shows the direction of the resultant image reported by the subjects in terms of a comparison stimulus.

Comparing Figs. 8 with 9, we can see that a resultant image of the test signal moves from its position in the noiseless condition to that of the direction opposite to the noise, in the same way it does when a real image is used as a signal. The amount of the angle shift in a sound localization increases along with the noise level relative to the signal level.

5. DISCUSSION

The sound image of a band noise shifted to the direction opposite to the noise source irrespective of the signal frequency and the kind of the signal, real or virtual. When the signal was a pure tone, the sound image shifted to the direction as long as the signal frequency was 1 kHz or below, and for the frequency of 2 kHz, the tendency of the subjects' perceptions varied from person to person. This difference can be explained by the fact that the one octave band noise centered at 2 kHz contains components down to 1.5 kHz.

It seems that a motion of localized direction is caused by a kind of (spatial) masking. This idea can be expressed as follows: Suppose localized sound has a certain extent in a psychological auditory space. This will be called an 'image', hereafter. Then the image of the signal and the noise overlap if they are close together. The overlapping part of the images can cause masking to lead to a deformation of the image of the signal. This deformation results in biasing of a 'perceived center' of the signal to an opposite side of the noise.

6. CONCLUSION

The effect of noise on sound localization for a band noise source was examined here. It is clarified that the sound image generally shifts to the direction opposite to that of noise.

Acknowledgement

This study was supported by the Sound Technology Promotion Foundation (STPF).

References

TRANIENT RESPONSE OF HUMAN HEARING SYSTEM

By Yoshimasa Sakurai
Dept., Engineering, Kansai Univ., 3-3-35, Yamate-cho, Suita, Osaka 564, Japan

Introduction
The reflection from a small rigid plane panel has a positive specular reflection and delayed large negative boundary wave\(^1\), which gives information about the size and location of the panel. Even if the panel is small, these two kinds of waves arrive with time intervals depending on their path differences. The reflected sound appears to be very weak. In order to explain this experience, we have to assume that a transient response lasts for a certain time on our hearing system. At the exposure of a positive rectangular pulse wave, the response, for instance, of a membrane shows a positive response that is followed by a negative transient. Our hearing system contains at least such a membrane as the ear drum. Then in order to simulate the system response by a transient that has a positive part followed by a negative reaction.

In the Visual Sound Field\(^2\) -- the spatial expression of impulse response in an auditorium, high order reflections do not give us much information in spite of their complexity. The calculated impulse response on the V.S.F. must be simplified by convolving the transient response of our hearing system to have more compact data for acoustical estimations of an auditorium.

1. Measurement of the Transient Response of Our Hearing System
The transient response of our hearing system was found by comparison measurements with the rectangular pulses of a width of 0.05 ms and an amplitude of about 107 dB, which was related to the level of the steady state sound pressure. Signals were generated by a loud speaker through a digitally corrected filter as shown in Fig.1. Two rectangular pulses were successively generated, with the above amplitude and width. They were not heard to be separated, but their loudness was greater than that of each pulse. Their loudness was compared with a reference rectangular pulse of a width of 0.05 ms which was generated about three seconds previously. The heights of the rectangular pulses were changed in five steps and compared with the loudness of two pulses of the same time interval. One of the five references signals was selected to have the same loudness as the two successive pulses. Half of the height of the reference signal gives a point on the transient response as shown in Fig.2(a). This procedure with different time intervals of two positive pulses found the thick curve of a transient response in Fig.2(a). When the sign of the latter pulse was changed to negative, the same comparison led to the thick curves in Fig.2(b).

Fig.2 Comparison measurements for obtaining a transient response for a rectangular pulse with a width of 0.05 ms.

The transient response thus obtained is shown in Fig.3(a). It is averaged over the results of four male students. The steep slope from the positive side to the negative side was obtained by connecting the point where two positive pulses were separately heard to the point where the sum of positive and negative pulse was larger than that of each pulse. The finishing point was assumed to represent the time when a positive and negative pulse were separately heard. The frequency characteristics of the response are shown in Fig.3(b) and (c). While the largest sensitivity of the steady state pure tone occurs at 3 kHz, it occurs about 300 Hz for the transient response.

Fig.3 Transient response of our hearing system for a rectangular pulse.

\(^{1}\) Reference signal
\(^{2}\) Spatial expression

![Diagram](image-url)
2. Linearity of the Transient Response

The ratio between the frequency characteristics of two rectangular pulses of widths of 0.05 ms and 0.07 ms was convolved with the response obtained by the pulses of 0.05 ms in Fig.3(a). The transient response thus calculated is similar to the measured response for the width of 0.07 ms as in Fig.4.

(a) Transient response  (b) Frequency characteristics of amplitude

Fig.4 Comparison of the response calculated from the result by 0.05 ms pulse and measured by 0.07 ms pulse.

3. Loudness Comparison of Reflections from a Large Rigid Panel and a Small One

A direct pulse from the loudspeaker and its reflection from a large or small rigid panel were examined by a listener at the seat in Fig.5(a). The centers and surfaces of the panels were located at the same positions. The direct and reflected pulses are shown in Fig.5(b) and (c). Because the loudspeaker faced equally to the center of a panel and a listener at the seat, precise rectangular pulses were not obtained. When four student

Fig.5 Loudness comparison on the reflection from a large and small rigid panel.

4. Discrimination of Two Rectangular Pulses

In order to find such an interference region in the concentric sphere, discrimination tests of two rectangular pulses from different directions were tried. Discrimination of the time interval for two successive positive rectangular pulse of a width of 0.05 ms generated from the front loudspeaker was 1.2 ms. When the latter pulse was produced by a different loudspeaker at 45° or 90°, it shortened to 1.15 ms or to 1.0 ms, respectively. The difference seems to be caused by the different head-related transfer functions. The directivity ratio between the frontal incidence and incidence at a different angle was transfer function or artificial head. Cross correlation function between the transient responses for frontal incidence and for an incident angle of 45° or 90°, was calculated by the measured directivity ratio. The discrimination time for two pulses for frontal incidence was 1.2 ms, where the cross correlation function was slightly on the negative side. Discrimination times for two other incident angles were obtained when the times have the same value on their cross correlation functions. Calculated times were 1.2 ms for 45° and 1.0 ms for 90° which observations correspond well to the measured discrimination times 1.15 ms and 1.0 ms, respectively. When a delayed pulse was negative, the discrimination time for frontal incidence was 2.5 ms. Cross correlation functions were also calculated and discrimination times were predicted to 2.4 ms at 45° and 90°. The measured values were 2.1 ms at 45° and 1.9 ms at 90°. This result different from the case with two positive pulses might have been caused by a rough approximation of the negative transient response. In this way, a discrimination time interval is very much affected by the head-related transfer function.

When transfer function of the head for the frontal incidence is $F(w)$, the directivity ratio of transfer function at a different incidence angle $\theta$ is $D_{\theta}(w)$, and the Fourier transform of a rectangular pulse of the width of 0.05 ms is $R(w)$, then the cross correlation function $\phi_{D}(\tau)$ between two directions is calculated by

$$\phi_{D}(\tau) = \int_{-\infty}^{\infty} R(w)F(w)\cdot R(w)F(w)\cdot D_{\theta}(w)\cdot \exp(j\omega \tau) \, dw$$

$$\phi_{D}(\tau) = \int_{-\infty}^{\infty} |R(w)|^2 |D_{\theta}(w)| \cdot \exp(j\omega \tau) \, dw$$

$\phi_{D}(\tau)$ is determined only by $D_{\theta}(w)$, which will be obtained by the measurements with an artificial head. Such a function $\phi_{D}(\tau)$ will give information to decide the interferable region in the concentric sphere on the V.S.F.

Acknowledgement

The author expresses his gratitude for the help of Mr. Hiroshi Morimoto, the engineer at Kanai University.

References

THE DEPENDENCE OF CRITICAL LEVEL ON TIME

W. Dixon Ward and Dorisann Henderson
Hearing Research Laboratory, University of Minnesota, 2530 University Ave. SE, Minneapolis, MN 55454, USA

The term "critical intensity" was first formally introduced by the field of noise-induced hearing damage by L. R. Hadd in 1954. He used it to describe the intensity of an exposure to white noise that just produced "pathological" instead of "physiological" auditory fatigue (Hadd, 1954a). However, for Hadd, pathological fatigue was a temporary threshold shift (TTS) that did not drop to zero by 15 min after termination of exposure; his studies with 2-min exposures apparently convinced him that there was more than a simple threshold of effect involved here, and that there was some sort of step function—a discontinuity—at the critical intensity (CI).

Although subsequent work has failed to show any such discontinuity in the relation between intensity and threshold for TTS, the general concept of the "maximum intensity or noise which the test subject can tolerate without pathological hearing loss" (Hadd, 1954b) has survived. After all, the cochlea is a biological system with seemingly fragile hair cells, supporting cells, membranes, and other structures that respond to static stimuli by vibration; with even greater amplitude as the intensity is raised, what is more likely than that this system has a break point, like a fine piece of crystal, beyond which permanent threshold shift (PTS) will be produced? In effect, whatever the truth, it is beyond our power to view the situation be likened to the straw that broke the camel's back? Given the complexity of the cochlear partition and our ignorance of the relative importance of the various morphological and chemical correlates of TTS and PTS, it is certainly not unreasonable to expect some discontinuity to exist.

CI for Impulses

In the ensuing third of a century, various forms of CI have been used to demonstrate the sharp discontinuity of effect associated with intensity that the term implies. For example, McRobert and Ward (1973) applied the term "critical level" (CL) to "the peak intensity level of an impulse of 0.5-μs A-duration, 20 mg which, delivered at a rate of one every 3 sec and by vibrations of 30 dB at 4 kHz", in a study that showed that if the intensity of the impulse were decreased from this CI (which had a median value of 152 dB SPL) by a factor of 10, then the ear could tolerate an indefinite number of impulses without developing any appreciable TTS. However, this particular "critical level" must be assumed to depend on all the parameters listed—the pulse duration, the number of pulses, and the rate of delivery, but none have shown the sharp discontinuity of effect associated with intensity that the term implies. For example, McRobert and Ward (1973) applied the term "critical level" (CL) to "the peak intensity level of an impulse of 0.5-μs A-duration, 20 mg which, delivered at a rate of one every 3 sec and by vibrations of 30 dB at 4 kHz", in a study that showed that if the intensity of the impulse were decreased from this CI (which had a median value of 152 dB SPL) by a factor of 10, then the ear could tolerate an indefinite number of impulses without developing any appreciable TTS. However, this particular "critical level" must be assumed to depend on all the parameters listed—the pulse duration, the number of pulses, and the rate of delivery, but none have shown the sharp discontinuity of effect associated with intensity that the term implies. For example, McRobert and Ward (1973) applied the term "critical level" (CL) to "the peak intensity level of an impulse of 0.5-μs A-duration, 20 mg which, delivered at a rate of one every 3 sec and by vibrations of 30 dB at 4 kHz", in a study that showed that if the intensity of the impulse were decreased from this CI (which had a median value of 152 dB SPL) by a factor of 10, then the ear could tolerate an indefinite number of impulses without developing any appreciable TTS. However, this particular "critical level" must be assumed to depend on all the parameters listed—the pulse duration, the number of pulses, and the rate of delivery, but none have shown the sharp discontinuity of effect associated with intensity that the term implies.
110 dB instead of 108 dB, only resulted in an increase of OHC destruction from 2.6% to 4.6%.

CI and Intermittence

At this point we were ready to return to the question of the CI. Intermittence reduces the effect of a given exposure, wouldn’t it therefore raise the CI? Specifically, would 120 dB still be above the critical level if the daily exposure were broken up into short bursts? Accordingly, a group of chinchillas was exposed to 120-dB noise, but only in 7.2-s bursts that followed the pattern described just above: the daily exposure, Monday through Friday, consisted of 10 of these 7.2-s bursts presented at 12-min intervals. We had intended to terminate the exposure after 5 days, at which time the cumulative exposure time would be 28 min; however, as these animals gave so little evidence, as indicated by behavioral audiometric changes, of any effect of the exposure, that we continued it for the full 9 weeks (45 days).

The histological analysis indicated that the CI had not been reached; the median OHC destruction was 5.3%, and in no animal did it exceed 10%.

This unequivocal result has implications both for the nature of the CI and for the equal-energy theory. First, the CI clearly depends on duration even for short bursts. Whether or not exposure led to the massive OHC destruction in the animals originally exposed for 22 min to 120 dB takes more than 7.2 sec to build up, and is largely recovered by the end of the 12-min intervening quiet interval, which at least raises questions as to the validity of the notion that simple mechanical stress can be the sole responsible agent involved. At any rate, if the human auditory system operates in a manner at all similar to that of the chinchilla, ceilings that are imposed on workers’ exposure levels regardless of duration, such as the 115 dBA of OSHA regulations (OSHA, 1983), will usually be incorrect, being overly protective for short bursts of noise.

The second implication of the results is that the equal-energy principle is not completely wrong, but merely needs qualification. The fact that the OHC destruction produced by 45 days of exposure to 107.2-s bursts of 120-dB noise at 12-min intervals was essentially the same as that produced by 45 days of exposure to 107 7.2-s bursts of 110-dB noise at 12-min intervals (5.3% and 4.6%, respectively) suggests that the equal-energy principle should be stated in this way: The permanent effect of a daily exposure is determined by the total energy of that exposure, provided that the temporal pattern of the exposure is held constant (and of course provided also that the critical level associated with any given noise burst that actually occurs is not exceeded).

Obviously, the relations among intensity, burst duration, and recovery interval that determine criticality in any but the simplest situation (a single uninterrupted exposure) are sure to be complicated. However, they must continue to be studied if we are ever to find a better way to summarize the hazardousness of a noise exposure in a single index that is more accurate than blindly integrating the sound energy entering the ear over the workday with no regard for temporal pattern.

Acknowledgment

This research was supported by NIH Grant 1-R01-NS-12125.

References


CONCEPTION D'UN SYSTÈME DE TRAITEMENT NUMÉRIQUE DES SIGNAUX SONORES TRANSITOIRES DE LONGUE DURÉE ET DE HAUT NIVEAU

M. Sawan, Y. Champoux, J. Nicolas
GAUS, Département de Génie Mécanique, Université de Sherbrooke, Sherbrooke, Québec, Canada

INTRODUCTION

Depuis plusieurs années, les applications qui nécessitent des ondes sonores transitoyes se multiplient. Au niveau de la mesure de la perte auditive provisoire due à des signaux impulsifs, on peut répertorier plus d'une douzaine de techniques de génération de signal. La grande majorité permettent ni de contrôler les paramètres importants ni de connaître les contenus temporel et fréquentiel. La mise au point d'un système composé de plusieurs filaments basés sur le microprocesseur et de plusieurs chaînes de son de haute puissance vient remplacer les oreilles à l'œuvre, les générateurs d'arc, les bruits industriels, etc, pour obtenir des signaux impulsifs de haut niveau (140 dB crête) grâce au haut-parleur JBL 2445J et de l'amplificateur de haute puissance (750 W). Autre chose d'autre, il s'agit de résoudre le problème de déformation des ondes générées par des méthodes de contrôle basées sur la compensation de la réponse en fréquence.

Les particularités du système sont: 1) la génération des signaux impulsionnels de haut niveau (140 dB crête) grâce au haut-parleur JBL 2445J et de l'amplificateur de haute puissance (750 W). 2) la durée des signaux peut dépasser 1 seconde grâce à l'amplificateur bicanaux 2032 de B&K qui permet l'acquisition des signaux longs (1-85 Echantillons), et à la technique de traitement (ZOOM-FPT).

La performance de technique de génération et de la souplesse de traitement ainsi que la répétitivité de signaux générée ont permis plusieurs applications dont la mesure de la perte auditive provisoire [4].

SYNTHESE DES MÉTHODES

Comme il a été montré auparavant [1,2], l'obtention d'une onde sonore p(t) semblable à un signal électrique théorique x(t) nécessite le calcul d'un signal corrige par:

\[ X_c(f) = \frac{X(f)}{H(f)} \times x(t) \]

où \( X_c(f), X(f) \) sont les transformées de Fourier des signaux corrige \( x_c(t) \) et théorique \( x(t) \),

\( H(f) \) représente la fonction de transfert de la chaîne de son utilisée.

\( k \) est un facteur d'amplification.

La méthode courante de contrôle qui a pour rôle d'atténuer les distorsions (bruit de fond superposé à l'entrée ou à la sortie de la chaîne) est basée à l'évaluation de la fonction de transfert à partir des auto-spectres et des spectres croisés de l'entrée et de la sortie de la chaîne. La figure 1 montre les résultats qui constituent le système complet, la tâche de contrôle s'effectue en deux étapes: 1) recherche de la réponse en fréquence, 2) calcul du signal corrige. Ces différentes méthodes ne sont valables que pour des systèmes linéaires, ou dans des zones linéaires de fonctionnement. En effet la génération des niveaux très élevés avait l'effet de détrôner la gamme dynamique des impulsions ce que nous avons incité à construire des nouveaux cornets des formes géométriques appropriées pour avoir le résultat visé.

Le traitement des signaux longs est composé de deux étapes: 1) Acquisition de l'entrée et de la sortie de la chaîne par la père rotative de l'analyseur bicanaux et transfert de ces signaux à l'ordinateur HP9816. 2) Calcul de H(f) par la nouvelle technique ZOOM-FPT [3] qu'il s'agit de décomposer un signal de N Echantillons en N blocs ayant chacun L Echantillons. L'effet de technique est la précision ou l'obtention des signaux M fois plus longue. La transformée du Fourier du signal est obtenue ainsi par:

\[ X(k) = \sum_{n=0}^{N-1} x(n) \cdot W_n^k \quad w_n = \exp (j2\pi n/N) \]

Posons que \( N = ML \), \( n = IM + m \) et \( k = RL + s \)

\[ X(l,s) = M^{-1} L^{-1} \sum_{m=0}^{L-1} \sum_{n=0}^{M-1} x(IM + m) \cdot w_n^{IM + m} \cdot w_s^{IM + m} \]

VALIDATION DE TECHNIQUE

La véracité du système permet un nombre illimité de signaux dans la bande 1-10 kHz de durée 0,02-1000 ms et de gamme dynamique dépassant 20 dB à un niveau crête de 140 dB. La figure 2 présente la gamme dynamique en fonction du niveau crête des trois cornets différents (a) et la gamme dynamique en fonction de la fréquence d'un seul cornet (b). L'effet du cornet au contrôle est montré à la figure 3.

La figure 4 montre la différence de précision due au traitement des signaux longs. Des tests ont montré le rôle inutile des adaptateurs en Y et de celui de la caisse de son commerciale lors d'ionection des ondes transitoyes.

La mesure de la perte auditive provisoire est une application qui nécessite le contrôle systématique de 15 paramètres importants (répétitivité, cadence, fréquences, spectres, niveau crête, etc), on peut voir un exemple des signaux utilisés à la figure 5 qui est la simulation d'un coup idéal de pistolet, ainsi que la figure 6 montre une impulsion et son spectre 1/3 octave plat.

CONCLUSION

Le traitement numérique des signaux et l'utilisation d'une chaîne de son de haute puissance assurent un choix illimité des ondes avec une excellente répétitivité. Cependant le contrôle des signaux longs et la construction des nouveaux cornets contribuent à la précision et à une gamme dynamique importante ce qui permet de réaliser des études systématiques soit:

- la perte auditive provisoire.
- la perte d'énergie par écarts.

REFERENCES


Figure 1: Schéma blocs du système de contrôle

Figure 2: Comportement des hauts-parleurs

Figure 3: Effet des cornets sur la gamme dynamique à 140 dB

Figure 4: Effet du traitement des signaux longs sur le contenu fréquentiel

Figure 5: Simulation d'un coup de pistolet

Figure 6: Simulation d'un spectre 1/3 octave plat
THE GROWTH OF TTS FROM IMPULSE NOISE

C. Laroche and R. Hétu

Ecole d'orthophonie et audiologie, Université de Montréal, P.O. Box 6128, Station A, Montréal, P.Q., H3C 3J7

INTRODUCTION

Contrary to steady-state noise, few studies have been conducted on the TTS growth curves in relation to the number of impulse noises of short duration. In fact, only three TTS studies on human subjects (1,2,3) have focused on this relation. It is generally stated that the growth of TTS is a linear function of the exposure time or the number of impulse noises (4). However, recent studies done on animals (5,6,7) and on humans (8) suggest a process which is neither linear nor logarithmic, but asymptotic. The goal of this study was to verify if this is the case for signals of short duration at peak pressures which induce relatively small amounts of TTS (< 15 dB).

According to the results from Ward et al. (1), the TTS growth curve in relation to the number of impulses is a linear function. The results from Loeb et al. (2) are not in agreement with this. A close analysis of their data shows a relation between TTS and the exposure time which varies according to the frequency: it is linear at 4 and 6 kHz, exponential at 8 kHz and apparently asymptotic at 12 kHz. Walker's results (3) suggest a function which is half-way between a linear and a logarithmic function. It is worth noting that these last results were re-analysed three years later using an asymptotic function, namely the Gompertz function (9).

The differences noted between these results could be attributed to many factors, such as the physical parameters used and b) in terms of the method used to analyse the data. These factors are described in Table 1.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Studies</th>
<th>Duration</th>
<th>Spectral content</th>
<th>Peak SPL</th>
<th>Number of exposures</th>
<th>Total Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ward et al.</td>
<td>A= 500 usec</td>
<td>centered at 1 kHz</td>
<td>140-155 dB</td>
<td>4</td>
<td>2-16 min</td>
</tr>
<tr>
<td></td>
<td>Loeb et al.</td>
<td>A= 40 usec</td>
<td>100 kHz</td>
<td>156 dB</td>
<td>3</td>
<td>500-800 sec</td>
</tr>
<tr>
<td></td>
<td>Walker</td>
<td>B= 20 usec</td>
<td>non speciff.</td>
<td>127 dB</td>
<td>3 or 4</td>
<td>500-800 sec</td>
</tr>
</tbody>
</table>

Comparing one study to another, it is obvious that all parameters differ. The spectral content alone or combined with another parameter, for example the peak pressure, could eventually explain the differences in the results. Furthermore, the severity of the exposure may not be comparable; in the case of Ward et al., the amounts of TTS induced were higher than in the other studies.

Concerning the analysis of the data, each author chose a different method: averaging over many frequencies, many subjects, many repetition rates and, in one case, over many subjects exposed to different peak pressures. It is possible that all this averaging mask a phenomenon which could be observed at individual level.

These data do not allow to define a rule which relates the exposure time to TTS for impulse noises.

Results obtained on animals have shown that an asymptote is reached after long exposure times. These studies (5,6,7) have corroborated, for 18 from 99 to 120 dB SPL and exposure times of 5 to 10 days, that TTS reaches an asymptote of 30 to 90 dB, depending on the condition, after an average of one to two hours of exposure.

This result has been confirmed on four human subjects, with an impact noise of 500 msec (8 duration). In less than 30 minutes, at a rate of 1 imp./sec., ATS of 4 to 18 dB have been measured (18=102-120 dB) (8).

The aim of this study was to verify if this phenomenon could be observed with short duration impulses in the same range of exposure times with human subjects (10).

METHOD

Three young adults with normal hearing have been exposed to 3 impulsive signals of short duration (8 duration: 1 to 3 msec, Lp=124 to 136 dB, rate=1 imp/sec., spectral content A=0.3-1 kHz, Br=0.3-3 kHz, C=0.3-4 kHz). These signals were produced by a computerized noise generator which permits a strict control of the physical parameters of the transient signal (11).

The maximal TTS has been limited to 15 dB at the most sensitive ear and frequency. The TTS were measured using the same method as in previous experiments (12). We determined the Lp which induced a target amount of TTS (or eventually ATS) of 8-12 dB after an exposure of about 24 minutes. Knowing the Lp, the experiment comprised 5 to 8 exposures lasting 4 to 48 minutes. Each exposure was separated by a recovery period of 20-24 hours. In all, the number of exposures and pre-exposure trials were about 50 per subject.

RESULTS AND DISCUSSION

The results for signal B (spectral content between 0.3 and 4 kHz) are shown in figure 1.

![Figure 1. TTS growth curves of subject 1 (Lp= 128, ATS= 4 dB), 2 (Lp= 127, ATS= 17.6 dB) and 3 (Lp= 125, ATS= 13.4 dB) for signal B.](attachment:image.png)
It is clear that TTS reaches a plateau after about 30 minutes of exposure for subject 3 (10^125 dB, ATs= 13.4 dB). For subject 2, the growth of TTS tends to stabilise, reaching a plateau in about 1 hour (10^127 dB, ATs= 17.6 dB), as estimated from the Guertler function. In the case of subject 1, we were faced with a variation which was unpredictable, uncontrollable and unexplained (i.e. 0 and 9 dB of TTS after 32 and 24 minutes of exposure respectively). Figure 2 represents the results for subject C (with spectral content similar to signal B). It can be stated that TTS is plateauing after about 30 minutes of exposure for subjects 1 (10^124 dB, ATs= 8 dB) and 3 (10^127 dB, ATs= 13.5 dB) and possibly after 1 hour for subject 2 (10^129 dB, ATs= 12.4 dB).

Figure 2. TTS growth curves of subject 1 (10^124, ATs= 8 dB), 2 (10^129, ATs= 12.4 dB) and 3 (10^127 dB, ATs= 13.5 dB) for signal C.

A plateau has also been noted for signal A (0.3-1 kHz) for 2 subjects but at TTS levels relatively low (<5 dB) and for 1 kHz 134-136 dB.

These results, as well as those from Hétu and Poirier (8) and from animals studies (5,6,7) demonstrate that the growth of TTS tends to stabilise within relatively short exposure times.

Comparing them with the data from earlier studies points out the need to study systematically the factors which determine the characteristics of the TTS growth curve as a function of exposure time to impulsive noise.

CONCLUSION

It has been shown in this study that when many distinct exposures and individual values are considered, the TTS growth curves tend to reach an asymptote. In these conditions, it is clear that studies on ATS, for impulsive noises as well as for continuous noises, deserve particular attention from researchers. Indeed, ATS provides reference against which comparisons among studies can be made (13).

This type of comparison is possible because:

1. a given noise induces a given effect with a steady value, i.e. independent of the exposure time (14);
2. the duration of exposure at the asymptote influences the recovery process (15);
3. the two preceding informations (ATS and recovery) are likely connected to the noisiness of noise in term of permanent damages to the hearing (13);
4. in term of patho-physiology, ATS allows to relate TTS to the risk of permanent hearing loss and damage as studied on the animals cochleas (16,17).

REFERENCES

EFFECT OF TEMPORAL PATTERN OF IMPULSIVE SOUNDS ON LOUDNESS

Sonoko Kuwano and Seichiro Namba

College of General Education, Osaka University
1-1 Machikaneyama, Toyonaka, Osaka, 560, Japan

1. INTRODUCTION

We have been conducting experiments on the loudness of various impulsive sounds. From the results of our previous experiments it was suggested that the loudness of impulsive sounds can be approximated by the total energy, i.e. A-weighted sound exposure level, LAE, as shown in Fig.1. However, precise examination of these results indicates that the temporal structure of stimuli has a significant effect on the loudness, though the effect is small. The effect of temporal structure was examined using non-steady state sounds (Fig.2-a), artificial impulsive sounds (Fig.2-b), and sounds with rise time (Fig.2-c). Even if the total energy was equal, the loudness varied according to the difference of temporal structures, as shown in Fig.3-a – c. A model of the dynamic characteristics of hearing was proposed on the basis of the results of these experiments. As shown in Fig.4, this model assumes overshoot at the onset of the stimulus, suppression in the middle, and after-effect at the cessation.

In our previous experiments the stimuli were presented monaurally, which means that the effect of masking on the result was considerable. When stimuli are presented dichotically, there is some central masking effect, but this is less significant than monaural masking. Therefore, in the present experiment the stimuli (shown in Fig.2-a) were presented dichotically in order to examine the effect of temporal structures in detail.

2. EXPERIMENT

2.1 Stimuli

Stimulus patterns are shown in Fig.5. The higher-intensity component of each standard stimulus (S0) was at a level of 83 dBA and 50 msec in duration (hatched section in Fig.5). The rest was at a level of 80 dBA. The total duration of the whole stimulus was 350 msec. An comparison stimulus (S1), steady state noise of 350 msec duration was used and its level was varied by 1 dB steps. White noise was used as a carrier.

2.2 Procedure

The higher-intensity component of each stimulus was presented to the left ear of the subject; the
rest of Ss and Sc were presented to the right ear. The method of limit was used. Subjects were asked to judge whether Sc was perceived as being louder or softer in comparison with Ss. There was an interval of 1 sec between Ss and Sc, and an interval of 2.5 sec between pairs. Four trials, two ascending and two descending series, were conducted for each stimulus pattern. The order of stimulus presentation was randomized with each subject.

2.3 Apparatus
White noise was generated by a noise generator (Brüel & Kjaer 1405). After regulating duration, interval, and sound level using the Programmable Sound Control System II, the stimuli were presented to the subjects through an amplifier (Yamaha CA XII) and headphones (THB 49) in a sound-proof room.

2.4 Subjects
Five subjects, two females and three males with normal hearing ability, participated in the experiment.

3. RESULTS AND DISCUSSION
The point of subjective equality (PSE) was calculated by averaging 20 judgments by 5 subjects for each stimulus pattern. The results are shown in Fig. 6. There is no significant difference in the loudness among the stimulus patterns except for pattern 7, which was judged as being softer than any other pattern. This result is quite different from those of our previous experiments, where patterns 1 and 7 were judged as being louder than the other patterns.

A comparison of the results of our previous experiments, in which the presentation of stimuli was monaural, with the results of the present experiment, in which it was dichotic, suggests the following interpretation. The phenomena observed in our previous experiments may be due to the effect of temporal masking. That is, when the higher-intensity component is located at the onset or at the cessation of the stimulus, backward or forward masking occurs. On the other hand, when the higher-intensity component is located in the middle of the stimulus, both backward and forward masking occur. This may be the cause of the overestimation of the loudness of patterns 1 and 7 compared with the other patterns.

In the present experiment, because the stimuli were presented dichotically, the effect of masking was much smaller. This may erase the differences in the perception of loudness when the seven stimulus patterns were used. The reason why pattern 7 was judged as being softer than the other patterns is not clear.

Other factors concerning temporal effects in the perception of loudness should be taken into consideration. Hashimoto et al. have presented data which suggest the existence of overshoot at the onset of a stimulus and after-effect at its cessation. Pastlić has proposed a model of dynamic hearing sensation. It may be the case that there is interaction among backward, forward, and simultaneous masking. Further experiments will be needed in order to differentiate various temporal effects and examine more closely the dynamic characteristics of hearing.

ACKNOWLEDGEMENT
The authors are grateful to Professor T. Kato of Otemon Gakuin University for his kind cooperation with this study. This study was supported by a Grant in Aid for Scientific Researches (G050100060), Ministry of Education.

REFERENCES
TOWARDS A HEARING AID WITH MULTICHANNEL AUTOMATIC GAIN CONTROL

R. Plomp, P.C. Anema and J.N. van Dijkhuizen

Department of Otolaryngology, Free University Hospital, P.O.Box 7037, 1007 MB Amsterdam, Netherlands

The introduction of very-large-scale integration in electronic circuitry opens new possibilities in hearing-aid design. This challenge to the hearing aid and to investigate in which ways its circuit can be improved in order to meet the demands of the hearing impaired as good as possible.

The authors of this paper are currently involved in the development and evaluation of a multichannel hearing aid with automatic adaptation of its amplitude-frequency response to the variable conditions of ambient sounds interfering with the speech signal. In this paper some basic considerations and experiments will be discussed.

BASIC CONSIDERATIONS

It may be convenient to discuss the requirements of a hearing aid, as far as relevant here, on the basis of a number of theses:

(1) The intelligibility of speech in noise should be adopted as the main criterion of the hearing-aid's performance.

Most hearing-impaired subjects complain about their difficulties in understanding speech interfered with other sounds. This is a consequence of the fact that, generally, the threshold elevation in quiet is accompanied by an elevation of the patient's threshold for speech reception in noise. The hearing aid is unable to improve the speech-to-noise ratio (at least for the most common case of one or more voices as the interfering sound; for a recent review on this topic, see Plomp, 1986). For this reason a hearing aid should affect the speech-reception threshold (SRT) in noise as least as possible, thus the presence of noise is the best condition to test the performance of an aid.

(2) The hearing aid must be provided with some system of dynamic-range compression.

The threshold elevation of the impaired ear in quiet is usually not accompanied by a ditto elevation of the sound-pressure level of uncomfortable loudness (UCL). For optimal speech understanding the signal should exceed the absolute hearing threshold over a frequency range as wide as possible but should not exceed the discomfort threshold. In order to have this condition fulfilled over a wide range of sound levels, the dynamic range has to be compressed.

(3) The amount of compression should be frequency-dependent, requiring a multichannel system.

Since noise goes very widely in practice, the compression of the dynamic range should not be the same for all frequencies, but should depend on the current noise spectrum. Without such a control, a strong low-frequency noise would automatically reduce the gain for the high-frequency components so that they become inaudible. This can be avoided by splitting the frequency range in a number of bands with separately controlled amplification.

(4) The amplitude-compression system should be linear for the speech signal.

Whereas the intelligibility of undisturbed speech is not very susceptible to nonlinear distortion such as amplitude limiting, the situation is quite different for speech interfered with other sounds. This is understandable since in the latter case the distortion results in intermodulation between the wanted and unwanted signals. Therefore, any system of dynamic-range compression has to be viewed, which means that it should be one or another version of automatic gain control (AGC) (for an excellent review of literature, see Braida et al., 1979).

(5) The automatic gain control system should be provided with relatively large attack and release times.

The main characteristics of any AGC system are the times required to adapt the amplification to a sudden increase and decrease of the input signal, respectively. Braida et al. (1979) defined the attack time (TA) as the time required for the output of a compressor to reach its asymptotic value within 2 dB after the input is increased suddenly by at least 20 dB, and the release time (TR) as the time required for the output to reach its asymptotic value within 2 dB after the input is decreasing by at least 20 dB. Large and small values of TA and TR result in a quite different behavior: in the case of large times (e.g., 500 msec) the gain adapts itself to the global levels of the speech sound and the noise, whereas in the case of small values (e.g., 20 msec) the gain follows rapid variations within the speech signal and the interfering sound.

In recent years the use of small compression times, particularly for TA, has become popular. The term 'syllabic compression' reveals that it is considered to reduce the dynamic range of the successive phonemes in current speech. There are, however, strong objections against the use of short times for TA and TR (cf. King and Martin, 1984) (a) the dynamic variations in speech are the carrier of information and any disturbance may affect intelligibility, and (b) short attack and release times will result in interaction effects, particularly if the interfering sound is characterized by rapid variations. Both arguments militates against small compression times. Moreover, it should be realized that the range between threshold level and UCL level exceeds for most hearing-impaired subjects the dynamic range of the speech signal, equal to about 30 dB so that in this respect small compression times are not needed.

Braida, Durlach et al. (1979) concluded in their review that the literature on syllabic compression was inconclusive and designed new experiments with a multichannel amplitude compression system (Lippmann, Braida, and Durlach, 1981). Their data confirm that, even under optimal compression conditions, small TA and TR values (160 msec and 20 msec, respectively) have a negative effect on the intelligibility of speech with small word-to-word level variations presented against a background of cafeteria noise. Other experiments confirmed this finding (e.g. Nabolék, 1983; Dreschler, Eberhardt, and Melk, 1984). More favorable results with a two-channel compression system (Laurence, Moore, and Glasberg, 1983; Moore, Laurence, and Wright, 1985) do not necessarily imply that the small TA and TR values applied were responsible for the improvement of speech intelligibility in noise; it seems probable that this improvement was due to the separate control of the two channels, similarly as was found by Ono, Kanazaki and Mizoi (1983) in their two-channel hearing-aid system.

EFFECT OF VARYING THE AMPLITUDE-FREQUENCY RESPONSE ON THE SPEECH-RECEPTION THRESHOLD IN NOISE

The use of multichannel automatic gain control implies that the amplitude-frequency response of the
hearing aid will depend on the frequency spectrum of the sound activating the microphone. As a consequence, in the case of a fluctuating background noise (for example, traffic noise), the speech spectrum will vary accordingly. We have to know the effect of these variations on speech intelligibility. This can be studied by measuring the speech reception threshold (SRT) for various frequency-response characteristics.

In a first experiment, the SRT was measured for a number of conditions with different slopes of the (stepwise) amplitude-frequency response. These characteristics were obtained by means of a series of 13 parallel 1/3-octave band-pass filters with center frequencies of 250, 315, 400, 500, ..., 4000 Hz, respectively. With this filter set slopes varying in 3-dB steps from -15 dB/octave to +12 dB/octave were realized. For these 10 conditions the SRT for sentences presented against a background of noise with the same spectrum as the long-term spectrum of the sentences was investigated with 20 normal-hearing listeners. (It is essential to realize that both in this and the second experiment the speech spectrum and the noise spectrum were shaped identically by the same set of band-pass filters.) The SRT (50% correct score) was measured monaurally by headphone with an adaptive procedure described elsewhere (Plomp and Mimpfen, 1979).

The average SRT values found are plotted in Fig. 1 as a function of the slope of the amplitude-frequency response. The diagram shows that, over a wide range from about -7 dB/octave to +10 dB/octave, our hearing is remarkably resistant against differences in this slope.

Whereas in the experiment just described the slope of the amplitude-frequency response was constant during listening to a sentence, dynamically varying slopes were studied in the second experiment. In this case a set of four parallel octave band-pass filters covering the range 250-6000 Hz were used. This circuit, programmed in a TMS-320-10 signal processor, was controlled by a PDP-11/10 computer in such a way that the gain factors of the filters varied nonsinusoidally along a dB scale. Eight different conditions were investigated: in half of them the slope varied between -5 and +5 dB/octave, in the other half between -10 and +10 dB/octave. The second parameter was the rate of the slope variations: 0.25, 0.5, 1 and 2 Hz. For these eight conditions the SRT for sentences in noise was determined for 20 normal-hearing listeners. The further details were the same as in the previous experiment.

![Fig. 1. SRT for sentences as a function of the slope of the amplitude-frequency response.](image)

![Fig. 2. SRT for sentences as a function of the rate of changing the slope of the amplitude-frequency response dynamically between -5 and +5 dB/oct (lower curve) or -10 and +10 dB/oct (upper curve).](image)

In Fig. 2 the effect of the dynamically varying slope of the response curve on the SRT is represented. It was verified that the SRT for the lowest rate of the slope varying between -5 and +5 dB/octave is equal to the SRT for a flat amplitude-frequency response. We may conclude that variations over a range from one extreme to the other of 10 dB/octave with a rate up to 1 Hz, corresponding to a transition within 0.5 sec, will have only a minor effect on speech intelligibility.

This research was supported by the Programmabureau Innovatie Hulpmiddelen Gehandicapten.

References


COMPARING AN 8-CHANNEL COMPRESSION HEARING AID WITH CONVENTIONAL AIDS IN SPEECH-BAND NOISE

E.W. Yund, H.J. Simon and R. Efron

VA Medical Center, 150 Muir Road, Martinez, CA, 94553, USA

INTRODUCTION

Recent work in our laboratory has been directed toward characterizing the suprathreshold auditory function of individuals with sensorineural hearing loss (SNHL) by means of a theoretical model of speech processing (Efron, 1977; Efron, Yund & Divenyi, 1979). This characterization has then been used to design, by computer simulation, a multichannel signal processing system to compensate for that particular individual’s hearing loss. The purpose of this report is to compare the performance of the model-defined “hearing aid” with that of an independently recommended conventional aid and with unaided performance. The action of all aids is simulated by means of digital signal processing software. Results from the first three hearing-impaired subjects tested using speech stimuli in speech-band noise will be presented.

METHOD

Subjects

The three subjects with cochlear pathology included two with normal-to-moderate low and mid-frequency hearing, abruptly sloping (greater than 20 dB per octave above 2 kHz) to moderate-to-moderately-severe levels in the high frequencies (S1 and S2) and one with a mild loss abruptly sloping to a moderately-severe level at 3 kHz and rising (S3). Subjects ranged in age from 34 to 64 years. Etiologies were unknown. However, subjects had a familial history of hearing loss with aging and two subjects were veterans with some noise exposure history. None of the subjects were hearing aid users at the time of these experiments.

Materials and Apparatus

The test materials consisted of eleven sub-tests of the closed-set CUNT Vowels and Syllable Test (NST) developed by Rosenk et al. (1975). The male voice stimuli, digitized from pre-recorded tape, were filtered into eight bands using the ILS signal processing package. The bands were six 1/2 octave bands from 750 to 4000 Hz in addition to one above 4000 Hz and one below 750 Hz. For the unprocessed condition, these bands were recombined after filtering while for the processed conditions the bands were selectively amplified before recombination. For the “conventional” aid, amplification within bands was a linear constant which varied across bands to match the frequency response of the recommended commercially available hearing aid. Peak clipping or compression was done after recombination to complete the “conventional” aid. For the “model” aid, amplification within bands was a function of the signal level determined individually for the subject and implemented in a way similar to Robinson and Huntington (1973). Speech spectrum noise (SSN), random noise shaped to the long-term average speech spectrum (French and Steinberg, 1987), was added to the speech signal prior to the processing.

Procedure

The 91 syllables of the NST were randomized, independently for each run, and presented monaurally to the subjects who responded to the 7 to 9 response fields for the sub-set appropriate for the particular stimulus.

Test signals were computer controlled and delivered through a Danavox SM-W wide-band receiver coupled to the subject’s earmold or a TDH-39 earphone in a circumaural cushion. All test items were presented monaurally to the ear judged to be the best ear for amplification.

All testing was done in a single-walled IAC booth. Sessions were two hours in length with rest periods between. Subjects to testing with stimuli in noise each subject ran at least 15 sessions without noise, unprocessed as well as processed. These quiet sessions served primarily as training—all aids tested worked well, and about equally so, in quiet.

There were three experimental conditions at each signal-to-noise ratio (S/N): (1) “unprocessed,” (2) “conventional” aid and (3) “model” aid. Two noise levels (60 & 70 dB SPL) each included five S/N ratios, +15, +10, +5, 0, and -5 dB. Within each S/N at each noise level, six runs of 91 syllables for each condition were done in random order; in this way learning effects, if any, were distributed equally across test conditions.

RESULT

Fig. 1 illustrates the subjects’ performance as measured in percent correct for each processing condition at each signal and noise combination. Percent correct scores are shown on the ordinate with the signal-to-noise ratios (S/N) on the abscissa. Each data point in the figure represents a mean of six repeats of the 91 item list for each subject at each S/N. The filled symbols represent the “unprocessed” condition. The open circles and X’s represent the “model” and “conventional” aids, respectively. The four panels of Fig. 1 show the data for subject S1 at 60 and 70 dB SPL SSN (upper two panels) and for S2 at 60 dB SPL SSN and S3 at 70dB SPL SSN (lower two panels). An analysis of variance was done on the raw data used to compute the percent scores of Fig. 1 individually for each subject and noise level. In each case, the analysis revealed significant main effects of processing condition (unprocessed, model aid, conventional aid; p<.0001) and, of course,
signal-to-noise ratio (+15, +10, +5, 0, -5 dB; p<.0001). The interaction of processing condition and S/N was significant at p<.01 for subject 31 with 70 dB SNR and at p<.0001 for the other three subject-noise combinations.

Tukey's HSD (Kirk, 1982) post hoc analyses performed for particular comparisons of interest, supported the following statements (significance levels are given only where all comparisons implied by that statement exceed that level): (1) The highest mean scores obtained with each noise level and processing condition were at the highest S/N levels (p<.01). (2) At 50 dB SPL SNR the highest mean scores were obtained for the model aid at every S/N (p<.01). With 70 dB SPL SNR, the model aid performed better than the conventional aid at lower S/N (-5 & 0 for 31 and -5, 0 & +5 dB for 33; p<.01) while the conventional aid performed better at the highest S/N (+15 for 31, +10 & +15 dB for S3; p<.01). (3) Both aids performed better than the unprocessed condition at all S/N for all subjects (p<.01), except for subject 37 at 50 dB SPL with 0 & -5 S/N, where the conventional aid and unprocessed condition were not significantly different—indeed, at 35 dB SPL both represent approximately chance performance (12%).

In summary, the model aid was always better than the unprocessed condition and was generally better than the conventional aid. The stimulus conditions where the conventional aid equalled or surpassed the model aid are relatively easy ones (higher intensities and S/N) while the stimulus conditions where the model aid excels are the most difficult ones (lower intensities and S/N).

DISCUSSION

The model used here to define the amplification function within the frequency bands of this 8-channel compression hearing aid was originally developed to account for dichotic pitch interactions in subjects with normal hearing. One stage of the model corresponds to the sound-intensity-to-neural-response transduction stage of the auditory system. The purpose of the present study was to determine the utility of this model, especially the intensity-response transduction stage, in hearing aid design. The model and the psychophysical methods used to measure the parameters of the model are described in detail elsewhere (Yund & Efron, 1977; Efron, Yund & Divenyi, 1979) only results concerning the performance of the model hearing aid will be discussed here.

The results obtained thus far indicate that the model aid is quite successful when compared with a well-fit commercially available (conventional) aid. Initial testing in quiet (data not presented here in detail) indicated little if any difference between the model aid and the conventional aid in the range of 40 to 90 dB SPL except for some subjects' comments that the conventional aid was very loud at the highest intensity. In the presence of speech band noise, the model aid was generally better than the conventional aid. Only at the highest intensity and S/N levels did the conventional aid produce higher speech recognition scores than the model aid, although results for S2 at 70 dB SNR differ from the other two subjects in this respect—model aid recognition scores were 81%, 74%, & 68% compared with conventional aid scores of 71%, 65%, & 58% for S/N of +15, +10 & +5 dB, respectively. One possible explanation for the relatively lower performance of the model aid with respect to the conventional aid in some subjects at high input levels (80 and 85 dB speech in 70 dB SNR) may be the conservative output intensity limit of the current version of the model aid: no band is amplified to produce an output level over 100 dB SPL.

Prose comparisons of our results with those of other studies of multichannel compression systems is difficult, but our results seem to be among the most successful in demonstrating an advantage for multichannel compression over conventional linear amplification with frequency-gain characteristics matched to the individual subject. The difficulties in comparing across studies begin with the choice of the "conventional" or "linear" amplification used. In our experiment, the conventional aid used for each subject was recommended by an independent outside audiologist based on a complete audiometric battery of tests and including speech recognition in noise with several possible aids. The "conventional" condition thus represents a modern, well-fit, commercially available hearing aid. Although such a "conventional" condition can be criticized as less than "optimal" for any particular set of stimuli, it is a realistic standard of comparison for any new multichannel compression aid throughout the full range of stimuli that hearing aid users encounter outside laboratory conditions. Such a realistic standard of comparison plus the inclusion of a range of signal and noise intensity conditions is essential in the evaluation of any multichannel compression system. (See the Introduction of Walker, Byrre, & Dillon (1984) for a more complete discussion of these issues.)

Laurence, Moore & Glaeser (1983) have demonstrated the superiority of a 2-channel compression aid over a good linear compression condition. Unfortunately, it is not possible to compare the relative success of their compression aid with ours due to the large number of differences in the two experiments. Nevertheless, their success reinforces our own conclusion that multichannel compression aids individually fit to the patient with neurotympanic hearing loss will improve speech intelligibility, especially in the presence of noise.

REFERENCES


Work supported by V.A. Rehabilitation R and D.
ÉTUDE ET MOdÉLISATION D'ÉCOUTEURS MINiATURES DE PROTHÈSES AUDITIVES.

J.P. Guilhot, J.L. Charles, C. Assais

L.A.M.I. 38 rue des 36 Ponts 31062 Toulouse Cédex
France

L'étude des écouteurs miniatures et de leurs charges acoustiques présente un intérêt particulier dans le domaine de la correction de l'audité : celle-ci fait largement appel à des prothèses électro-acoustiques dont le gain final est constitué d'un écouteur miniature débitant dans une charge acoustique adaptée à la morphologie du sujet appareillé.

La miniaturisation de ces écouteurs permet la réalisation d'appareillages discrets, mais rend quasi impossible une mesure acoustique directe de la courbe de réponse de l'appareillage in situ. On convient dès lors l'utilité d'un modèle d'écouteur et de charge permettant la prévision de la courbe de réponse d'un écouteur débitant dans une charge de géométrie connue.

MOdÈLE D'ÉCOUTEUR

La modélisation directe des écouteurs miniatures n'est pas envisageable, et la solution utilisée consiste à modéliser l'écouteur par un quadripôle électro-acoustique dont les quatre coefficients $A$, $B$, $C$ et $D$ sont des nombres complexes fonction de la fréquence. $E$, $T$ et $L$ sont la tension et le courant aux bornes de l'écouteur, $P$ et $Q$ la pression et le flux de vitesse acoustique dans la charge :

$$
\begin{pmatrix}
P_1 \\
Q_1 \\
\end{pmatrix} =
\begin{pmatrix}
ch(jkL) & Z_s(hjL) \\
Z_s^* & ch(jkL) \\
\end{pmatrix}
\begin{pmatrix}
P_2 \\
Q_2 \\
\end{pmatrix}
$$

où $k$ et $Z_s$ sont des nombres complexes fonction de la fréquence, de la géométrie du conduit et des paramètres visco-thermiques de l'air. En particulier, la partie imaginaire du nombre d'onde $k$ est liée à l'atténuation de l'onde due aux effets visco-thermiques.

Dans le cas général, les charges sont constituées d'une série d'éléments de conduits circulaires de géométries différentes, et le modèle résultant s'obtient par la mise en cascade des quadripôles correspondant à chaque élément. La matrice de chaîne du modèle est alors le produit des matrice de chaîne de chaque quadripôle.

La connaissance de la condition de fermeture de la charge (air libre, microphone de mesure) donne également lieu à une modélisation appropriée et permet alors de donner l'expression de l'impédance acoustique de la charge complète.

MISE EN ŒUVRE DE LA MÉTHODE DES 2 CHARGES

La mise en œuvre de la méthode des 2 charges a été envisagée pour un modèle d'écouteur miniature de prothèse auditive. Les 2 charges retenues ont été construites de façon à obtenir des impédances acoustiques nettement différentes sur le domaine fréquentiel d'étude 100, 10000 Hz, afin d'obtenir les meilleurs résultats possibles. Le dispositif expérimental permettant de mesurer la réponse des grandeurs électriques aux bornes de l'écouteur, et acoustiques dans chacune des charges, a été appelé à un analyseur de Fourier bicanal piloté par un ordinateur sur lequel un programme a été écrit pour assurer le déroulement automatique de l'expérimentation. Les résultats expérimentaux ainsi obtenus sont alors utilisés par un programme de calcul incluant la description du modèle de charge, ce qui nous a permis d'obtenir les valeurs de chacun des coefficients complexes $A$, $B$, $C$ et $D$ du quadripôle associé à l'écouteur étudié. Au total, c'est un fichier de 1600 valeurs réelles qui est stocké sur le disque dur de l'ordinateur, correspondant à un pas en fréquence de 50 Hz.
APPLICATION DES MODÈLES

L'une des principales applications de ces modèles est la prévision de la courbe de réponse d'un écouteur miniature dans une charge de géométrie connue. Des exemples de résultats obtenus sont illustrés par les figures 2 et 3, où l'on a retenu des cas de géométries de charges couramment rencontrées avec les écouteurs de prothèses auditives. Les résultats prévus par les modèles d'écouteur et de charge figurent en trait interrompu, et ceux obtenus expérimentalement en trait plein.

Fig 2: courbe de réponse d'un écouteur miniature prévue et expérimentale, pour la géométrie de charge indiquée.

Fig 3: courbe de réponse d'un écouteur miniature prévue et expérimentale, pour la géométrie de charge indiquée.

L'examen des résultats obtenus par modélisation et expérimentalement montre une bonne coincidence (dans une fourchette de 3 dB), ce qui traduit la fiabilité des modèles décrits dans cette étude.

CONCLUSION

L'étude théorique des écouteurs miniatures et des charges acoustiques qui leur sont associées nous a permis de développer des modèles leur correspondant sous forme de quadripôles électro-acoustiques et acoustiques.

La validation expérimentale de ces modèles montre leur fiabilité dans la prévision du comportement acoustique des écouteurs miniatures et de leur charges. En particulier, ces modèles permettent la prévision du niveau de pression acoustique développé par un écouteur miniature dans une charge quelconque de géométrie connue, ce qui constitue un élément de réponse aux problèmes métrologiques liés à l'utilisation des prothèses auditives.

BIBLIOGRAPHIE


A new method of loudness category scaling has been developed and tested at the University of Würzburg on more than 1000 ears. Originally we used this test to evaluate hearing aids. A much wider application though is in testing loudspeakers or headphones. Loudness is a function of frequency and intensity, both parameters are very important for transducer evaluation. The speed and accuracy of this test is superior to free field reference tests. We will show data of headphone tests and compare the results to other methods. The complete test is recorded on a CD Laser disc, produced by WESTRA GmbH of Germany and simple to operate. A short description of the test procedure:

1. The subject is instructed that it will be exposed to two short bursts of 1/3-octave filtered noise of two seconds duration, two seconds pause and another two seconds of exposure. The intensity of the noise is random between 30 dBA SPL and 90 dBA SPL. The two bursts are of the same intensity.

2. Then the subject is asked to scale the intensity or loudness into one of the five categories: very soft, soft, medium loud, loud, very loud. This scale is presented to the subject in the following form:

<table>
<thead>
<tr>
<th>0</th>
<th>not heard</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>very soft</td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>soft</td>
</tr>
<tr>
<td>16</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>medium loud</td>
</tr>
<tr>
<td>26</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

The subjects response is a number between 0 and 50. This means that each of the five categories can be individually divided into 10 additional steps.

3. Between 9 and 16 random intensities are presented and scaled. It takes between 0 and 10 minutes to scale one frequency. Usually we do not test a single subject for more than 30 minutes during one session. Depending on the accuracy required we need about 10 to 20 subjects to describe a particular transducer.

4. A regression function is now calculated and if a single response is not within a predefined error range an automatic recheck is done, i.e. more samples are presented.

5. The calculated regression curve is now the subjects loudness function.

6. We can do this test at different center frequencies. The WESTRA CD-disc has 24 different ones, from 63Hz to 12500Hz.

7. We can average the responses of a group of test subjects and subtract this mean response at every single frequency from normal responses we averaged in a free sound field.

8. This freefield response can easily be recalculated into a diffusefield response.

We will present diagrams of scaled frequency response curves.
Zur Problematik


Besonders der Effekt der unterschiedlichen Lautheitsempfindung ist in vielen Jahrzehnten immer wieder Gegenstand wissenschaftlicher Untersuchungen gewesen (1,2,4,5).


Messaufbau und Versuchsdurchführung


Der Umschalter S₁ legte das Signal zum einen auf den Lautsprecherschallweg mit dem Verstärker A₁ und dem Lautsprecher L₁ und zum anderen auf den Kopfhörerweg mit dem Verstärker A₂, von dem aus binaural aber monophon der Kopfhörer H versorgt wurde.


Wenn nach mehrmaligen Umschalten im Vergleich zwi- schen Lautsprechereingriff und Kopfhörereingriff ein Lautheitsunterschied mehr festgestellt werden konnte, wurde durch Drücken der Taste S₄ im Schalter S₃ der mit S₄ eingestellte Lautstärkepegel A₄ aufge- speichert.

Die Werte wurden für acht Oktaven zwischen 100 Hz und 12,5 kHz für die vier Schalldruck-pegel 60, 70, 80 und 90 dB in einem von den Proban- den viertel registriert. Die Lautsprecheransteue- rung erfolgte dabei jeweils so, daß die abgegebene Schallqualität vom Rechner über die Treiber A₁ und A₂, entsprechend dem Oktavekasten des Testsigs- nals eingestellt wurde. Jede von vier durchgeführten Versuchsergebnissen mit beiden Signalschleifen. Jeder Meßdurchgang in 32 Ver gleichungen (8 Oka- 

Versuchsergebnisse

Wie vorherergehende Messungen zeigte auch der Einsatz eines Musiksignals stark personenbezogene Ergebnisse. Von der Mittelwerte über die Schalldruckpegel auf unterschiedlichen Probanden durfte daher zunächst abgesehen, um schwach ausgeprägte Effekte nicht sofort zu ver- decken. Der Vergleich der vier Messwerte zeigte, daß auch bei dem ungewohnten Signal eine gute Reproduzierbarkeit bis auf wenige Ausnahmen bewertete sich die Standardabweichung deutlich unterhalb 2,4 dB. Die Mittelwertbildung nach jedem Durchlauf erreichte sehr schnell stabile Werte.

Schlußfolgerungen und Ausblicke


Die bisher festgehaltenen Werte lassen noch keine endgültigen Schlußfolgerungen zu, da die Ausprägung der Effekte nicht übermäßig von den Reproduzierbarkeitsleitdränzen entfernt liegt. In weiteren Messreihen soll ermittelt werden, ob durch Veränderung der Hörentfernung zum Lautsprecher (sie betrug in den vorliegenden Messungen 2 m) sich Unterschiede einstellen und ob sich weitere Abweichungen durch Übergang auf stereophone Darbietung ergeben.

Literatur:
(1) C.M.B. Anderson und L.S. Whittle, Physiological Noise and the Missing 6 dB, Acustica Vol. 21 (1971), S. 261...272
(2) Leo L. Beranek, Acoustic Measurements, New York, London 1948, S. 730...743
(5) Günther Theile, Untersuchungen zur Standardisierung eines Studioskopfhörers, Rundfunktechnische Mitteilungen Heft 1, 1983, S. 17...26
(6) E. Werner, Headphones - still a lot of questions, AES Preprint 2272, 77th Convention in Hamburg 1985
(7) E. Werner, Zur Dynamik des Hörens mit Kopfhörern, NFG-Fachtagung "Hörkundefunk", Mannheim 1985
Trotz hochpräziser physikalischer Meßmethoden ist die Suche nach einem geeigneten und aussagekräftigen Meßverfahren für Kopfhörer noch nicht abgeschlossen. Im Unterschied zu Lautsprechern kann beim Kopfhörer nicht als physikalische Zweiwegrinne auf ein Schallfeld in einem reproduzierbar zu beschreibenden Wiedergaberaum zurückgegriffen werden. Das von Kopfhörer erzeugte Schallfeld ist zwar im Gehörgang des Probanden mit der Ohrröhre verbunden, jedoch ist die Neigung und Qualität solcher Art gewonnener Ergebnisse äußerst problematisch.

Die IEC-Norm 268-7 (3) sieht daher als Meßverfahren für Kopfhörer einen Lautheitsvergleich mit einem Lautsprecher vor. Durch diesen Lautheitsvergleich ist der Mensch aktiv in das Meßergebnis einbezogen.


Meßaufbau zur Einstellung des Mittendrucks


Meßablauf


Um Einstellungssicherheit zu erbringen, werden die Messungen bei jeder Versuchsperson viertelmal wiederholt, eine Prozeß, die bei der Zeitbedarf von etwa fünf Minuten für 20 Personen sehr schnell erledigt ist.

Meßergebnisse


Schlußfolgerungen und Ausblicke

Das Verfahren erwies sich als äußerst schnell und bequem reproduzierbar als Meßablauf, bei denen der Kopfhörer immer wieder abgesetzt werden muß. Während derartigen Verfahren Streuungen in der Größenordnung von vorzugsweise 2 dB liefern, liegt die Mehrzahl aller Streuwerte bei Verwendung des Mittendrucks um 1 dB und darunter.

Eine Reihe von Verbesserungen ist für die Messrichtung vorgesehen. So wird ein Kopfbügel aufgebaut, der eine Vielzahl unterschiedlich konstruktiv gestalteter Kopfhörerkapseln zu befestigen erlaubt. Er soll weiterhin den Andruck der Kopfhörerkapseln entsprechend der Verwendung des Originalbügels einstellbar erlauben.

Etwas störend ist beim beschriebenen Meßaufbau, daß sich neben der Position des Schallsignals während der Veränderungen der Tellerstellungen auch die Gesamtlautheitsumeindruck verändert. Durch einen geplanten und abgestimmten Teller im Referenzweg läßt sich die Gesamtschallleistung so beeinflussen, daß auch die Summenlautheit annähernd konstant bleibt.

Die ersten der mit dieser Methode ermittelten Übertragungsfunktionen mit denen nach dem Lautheitsvergleich müssen in größerer Zahl noch durchgeführt werden.

Literatur:

1. C.M.B. Anderson and L.S. Whittle
   Physiological Noise and the Missing 6 dB
   Acustica Vol. 21 (1971), S. 261...272

2. Leo L. Baranek
   Acoustic Measurements
   New York, London 1979, S. 730...739

3. IEC Publication 268
   Sound System Equipment
   Part 7: Headphones and Headsets
   Genf 1983

4. Wayne Rudden
   The case of the missing 6 dB
   J. Acoust. Soc. Am. 73(3), March 1982,
   S. 650...659

5. Floyd E. Toole
   The Acoustics and Psychoacoustics
   of Headphones
   2nd AES International Conference
   The Art and Technology of Recording
   Anaheim 1984, Preprint C1006

6. E. Werner
   Zur Dynamik des Hörers mit Kopfhörern
   NTG-Fachtagung "Hörrundfunk"
   Mannheim 1985

Fig. 1 : Measurement Setup
Bild 1 : Meßaufbau

Fig. 2 : Values of 4 measurement cycles
Bild 2 : Ergebnis von 4 Mesdurchgängen

Fig. 3 : Results with test object
Bild 3 : Ergebnis mit Testkapsel
NOTES ON DUMMY HEAD MICROPHONE FOR HEAD RELATED STEREOPHONY (OS8)

T. Miura, K. Okabe and H. Hamada
Tokyo Denki University
2-2 Kanda Nishiki-cho, Chiyoda-ku, Tokyo 101, Japan

1. INTRODUCTION

Head related stereophony (HRST) using a dummy head microphone (or head and torso simulator: HATS) is a powerful tool to study acoustics of concert halls and listening rooms [1]. Miura [2] proposed a standardized version of HRST called orthostereophonic system (OS8).

If the HRST system performance is perfect, a listener perceives just the same auditory impressions as he is in original sound fields (Fig.1). However, when HATS and reproducing system are not optimum for the listener, the original sound fields cannot be reproduced and he perceives erroneous localization azimuth shift and degradation of sound images. Furthermore, if the system is less perfect, so-called "forward-to-backward confusion" or "in-head-localization" arise.

In the case that the HRST system contains two reproduction loudspeakers and equalization digital filters, such as our OSS, these defective auditory impression can be minimize by (1) optimizing HATS for the listeners and (2) optimizing digital filters for the system.

Previous paper [3] and re-examined report [4] describe the physical performance of OSS and its effects on localization in horizontal plane. In this article, we discuss the possibilities of the standardized OSS for HRST. Its treatment is based on the relationship between localization in horizontal plane and interaural differences (ID).

2. APPROACH

Previous paper and report [3,4] describe the psychometrical relations between ID variations and horizontal localization shifts or degradations of perceived images. We define the ID variations as discrepancies of interaural time and intensity difference (ITD and IID) in simulated sound fields from those in original sound fields.

2.1 Localization azimuth shift in horizontal plane

Localization azimuth shift $\Delta \theta$ denotes the difference of the localization azimuth in the simulated field to that in the original field. We formulate $\Delta \theta$ as eq.(1).

$$\Delta \theta = \frac{(\Delta T_s - \Delta T_o) + R(\Delta P_s - \Delta P_o)}{aR + K} \quad (1)$$

where terms $\Delta T_s - \Delta T_o$ and $\Delta P_s - \Delta P_o$ denote the ITD variation and the IID variation respectively and $R$ means time-intensity trading ratio. $\Delta T_s - \Delta T_o$ is a mean value below 1 kHz and $\Delta P_s - \Delta P_o$ is averaged with 1/f weightings below 1 kHz. $K$ and $a$ are constant (see ref.[3] for derivations).

From the experimental results [4] on 5 subjects in Fig.2, $\Delta \theta$ in degrees was correlated to the perceived localization shift $\Delta \theta_m$ by eq.(2).

$$\Delta \theta_m = 0.585 \Delta \theta - 0.24 \quad (\text{deg}) \quad (2)$$

Eq.(1) and (2) restrict the tolerance limit of the ITD and IID variations.

2.2 Degradation of perceived sound image

If the HRST system is imperfect, listeners perceive broadened images. The degradation of perceived sound images $\Delta \theta$ can be modeled as eq.(3):

$$\Delta \theta = \sqrt{(\Delta T_s - \Delta T_o)^2 + (|\Delta P_s - \Delta P_o| C_{p})^2} \quad (3)$$

where $|\Delta T_s - \Delta T_o|$ denotes the absolute value of the interaural phase difference (IPD) variations averaged below 1 kHz. On the other hand, $|\Delta P_s - \Delta P_o|$ is the averaged absolute value with 1/f weightings below 16 kHz. $C_p$ and $C_p$ mean weighting functions for

![Fig.1 Construction of Orthostereophonic system.](image)

![Fig.2 Relations between perceived localization shift in horizontal plane $\Delta \theta_m$ and calculated $\Delta \theta$ from interaural difference (ID) variations.](image)
IPD and IID variations respectively. We adopted the inverse of interaural just-noticeable-difference (JND) as Cφ and Cp. Cφ was calculated using the JND data of IPD by Yost [5].

Eq. (3) suggests that a "norm" in a metric space such as the IPD-IID variations plane can be directly related to the degrees of the perceived image degradations (also see ref. [6]).

Results on 5 subjects in Fig. 3 show that ΔQc is a good index of the subjective image degradations when assuming the direct proportion between subjective degradations and variances of the localization azimuth. That is, we assume subjects' reports on the localization azimuth fluctuate as the perceived images broaden. We denote ΔQm as the ratio of the standard deviation of the localization azimuth in the simulated sound field and those in the original sound field. As expected, ΔQm in Fig. 3 converges to 1 as the ID variations become small.

Results on rating of the perceived images using method of successive categories suggest that listeners perceive the "unnatural" broadened image when ΔQc exceeds 8. We can conclude, therefore, ΔQc must be less than 8 for the perceived image degradation.

Fig. 3 Relations between degradation of the perceived image ΔQm and calculated ΔQc from ID variations.

2.3 Calculation of ID variations

As a first approximation, we modeled a head as a sphere and calculated ID variations because the effects of pinna are negligible for the frequencies below 1 kHz. The reference head has a radius of 7.94 cm, which agrees with mean head width of Japanese male adults. We calculated the ID variations of a small head (r=7.356 cm) and a large head (r=8.524 cm) relative to the reference head. These radius values were chosen to cover 90% range of the population.

3. RESULTS AND DISCUSSIONS

Fig. 4 shows the calculated results of the ID variations as a function of sound source azimuth. The broken lines indicate the tolerance limits of the ID and the ID variations estimated by using Mills's MAA (minimum audible angle) [7] and Eq. (1) and (2). The results suggest that a HATS with average dimensions is useful because even the listeners with small or large head cannot detect localization azimuth shifts when listening to the materials recorded by the HATS with average head dimensions. Additional measurements of the ID variations on 20 subjects support our results.

From the measurement results, we also calculated ΔQc for each subject using the average IPD and IID of all subjects. All the ΔQc did not exceed 5, which means the degradation of perceived sound images still be small enough when listening to materials recorded by the HATS with the average ID.

Fig. 4 Tolerance limits derived from Mills's minimum audible angle (MAA) and and eq. (1) and (2). ID variations as a function of sound source azimuth were calculated by a spherical head when its radius is small (r=7.356 cm: ●) and large (r=8.524 cm: ○) relative to the average (r=7.94 cm). These values were chosen to cover 90% range of the population.

REFERENCES

IMPROVING THE SPEECH-INTelligIBILITY BY TELEPHONING
in a Noisy Environment

H. W. Gierlich

Institut f. EI. Nachrichtentechnik der RWTH Aachen
Meilenter Str. 23, D-5100 Aachen, Western Germany

1. Introduction

During telephoning different noise-signals reach the ear of the listener. Distortion, produced on the line will not be discussed in the following. More annoying are noise-signals in the room of the listener which reach the ear of the listener in different ways (1/1, 2/2). Using a normal telephone, you have an electrical sidetone-path - that means that you receive room noise picked up by your microphone in the handset and reproduced by the loudspeaker at your ear- and an acoustical sidetone path - that means that room noise reaches the ear of the listener by the "acoustical leak" between the ear and the handset. You also receive room noise by bone-conduction and by the other ear, which is not covered by the handset. That the noise-signal reaches this ear directly without any filtering. The transfer function of the acoustical leak is a function of the pressure applied to the handset and the individual ear geometry. The transfer function of the electrical sidetone is a function of the complex line-impedance. It could be shown, that sidetone by bone conduction is not relevant so that you can ignore this sidetone-path. To evaluate the influence of the remaining sidetone-paths, their exact transfer functions have to be known. Then the influence of the different sidetone-paths on the speech-intelligibility can be investigated by hearing-tests.

2. Measuring the Different Transfer Functions

For measuring the transfer functions of the acoustical sidetone-path and the signal-path, a special helmet was constructed. With this helmet it was possible, to place the handset in a defined position relative to the ear of the listener. Using this construction, different constant pressure forces between the handset and the ear of the listener could be applied to the ear of the listener. The measurements were made in a frequency-range from 200 Hz up to 5 kHz with 6 different test persons. The pressure forces were 13N, 8N and, for it is a problem (compliance of the outer ear), to define a pressure below 8N, at distances of 2mm, 4mm, 6mm and 8mm related to 8N pressure. Measurements were conducted by a miniature electret microphone which was positioned in the ear canal of the different test persons. In Figure 1 you see the transfer function $|H_{EP}(f,F)|$ (f-frequency, F-force of pressure) of the earpiece averaged over 6 test persons at several pressures, measured in an anechoic chamber. The Figure shows that with increasing pressure lower frequencies are better transmitted to the ear channel. At a pressure of 8N, the transfer function is nearly constant from 500Hz to 2kHz. Higher pressures than 13N are not realistic.

The transfer functions of the "acoustical leakage" $|H_{AL}(f,F)|$ are shown in figure 2. They were measured in a diffuse sound field using white noise. In the interesting frequency range the transfer function is similar to a low-pass filter. The cut off frequency and the slope of decrease are governed by the pressure forces. An increase of pressure force applied to the handset causes a better attenuation of frequencies above 400Hz. But at 8N pressure a maximum attenuation is reached. A higher pressure produces no better attenuation because the deformation of the outer ear is limited and an acoustical leakage remains.

![Fig. 1: Averaged transfer functions from ear-piece to the ear of the listener](image1)

![Fig. 2: Averaged transfer functions of the "acoustical leakage"](image2)

3. Simulation and Hearing-Tests

With knowledge of all relevant transfer-functions it is possible, to simulate telephoning in a noisy environment. The arrangement to simulate telephoning in a noisy environment is shown in Fig. 3. Using an artificial head (4/4) it is possible, to record both ear signals in different environments. A microphone is placed at the mouth where normally the microphone of the handset is positioned. The microphone-signal is filtered by the transfer function of a telephone-microphone $H_{FS}$, by the transfer function of the electrical sidetone-path $H_{FS}$ and by the...
transmission function $H_{EP}$ between earpiece and ear. The left ear signal of the dummy head is filtered by the transmission function $H_{AL}$ of the acoustical leakage. The right ear signal reaches the ear of the listener without filtering. The speech signal, monosyllabic words, is filtered by the transfer function of the telephone microphone by the transfer-function of the telephone line $H_{L}$, and by the transfer function $H_{EP}$. The different signals may be individually attenuated so that in the simulator the left ear signal is generated as follows:

$$P_{BE-1}(f,f,l) = k_1P_{BE}(f,f) + k_2P_{BE}(f,f) + k_3P_{BE}(f,f) + k_4P_{BE}(f,f)$$

The right ear signal is generated to

$$P_{BE-2}(f) = k_4P_{BE-2}(f)$$

Fig. 3: Simulation of telephoning in a noisy environment

Using headphones, test persons are able to hear as if they had really telephoned in the room. The signal $s(t)$ was a rhyme-test (/S/) in German language. The noise signals used in the room were white noise and noise with a power density spectrum similar to the power density spectrum of German speech. Some results of this test are shown in Fig. 4 where the speech-intelligibility in % depending on the signal to noise ratio for two different line-simulations is shown. In Fig. 4a a line, consisting of 4km with 0.6mm diameter and 2km with a diameter of 0.6mm, between the two telephones was simulated, in Fig. 4b the simulated line-length was 0km. In both cases first the electrical sidetone-path (1) and then the acoustical sidetone-path (2) was switched off. Switching off the acoustical sidetone, speech-intelligibility increases 10-20% at the same signal to noise ratio. Switching off the electrical sidetone-path you get an improvement of 20% up to 60%. The fourth graph (1) (Fig. 4b) shows the results when switching off the noise signal of the right ear. It is obvious, that speech-intelligibility in a diffuse noise field can not be improved by switching off this signal. In this case only the signal of the left ear is analysed (see /6/). In the fifth graph (1) speech-intelligibility is shown when attenuating the signal of the electrical sidetone-path by 10dB and amplifying the signals transmitted by the earpiece by 5dB. This experiment shows, that when using an adaptive hybrid or a voice switched attenuator at the microphone and an amplifier at the earpiece, speech-intelligibility increases by 30-40%.

Fig. 4: Speech-intelligibility for 2 different line-simulations

1. normal situation
2. acoust. sidetone off
3. right ear signal off
4. left ear signal off
5. see text

4. Conclusion

Hearing tests showed, that no improvement can be obtained (in a diffuse noise field) by suppressing the noise signal at the ear which is not covered by the handset. To improve speech intelligibility it is better to suppress the signal of the electrical sidetone-path using a voice switched attenuator/amplifier or an adaptive hybrid. If the electrical sidetone is reduced, suppressing the acoustical sidetone can be achieved by using a special sealing ring similar to the ones used for head phones, by amplifying or compressing the signal. These investigations result from a projekt supported by the german company TELEFONICA (TEN).

3/ Genuin, K.: "Ein Geräuschdiagnosesystem zur Analyse von Schallereignissen unter Ausnutzung des Nachrichtenempfängers Menschliches Gehirn" FASE/DAGA 82, Berichtsband, S.443-446
OCCUPATIONAL DEAFNESS: WHAT WAS LEARNED DURING THE ERA OF SYSTEMATIC, NON-ELECTRONIC INQUIRY, 1875-1925

G R C Atherley and W Noble

Canadian Centre for Occupational Health and Safety, 250 Main Street East, Hamilton, Ontario, L8N 1H6, Canada

The University of New England, Department of Psychology, Armidale, NSW 2351, Australia

Introduction

From 1875 onward, as discussed more fully by Atherley and Noble (1986), we see the emergence of systematic investigations into deafness among industrial workers.

In the present paper, we focus on the reports of Gottstein and Kayser (1881) and of Barr (1886). Their use of occupation-based epidemiology challenges contemporary orthodoxy. Barr’s use of self-report assessment of hearing handicap addresses an issue still live in the evaluation of industrial deafness.

Gottstein and Kayser's Study: Occupation-Based Epidemiology

Gottstein and Kayser's report represents a methodologic advance. For the first time, so far as we can tell, the hearing of the target group, blacksmiths and metal workers, is compared with that of a control group, bricklayers, matched in certain important ways with the target sample, but unexposed to loud noise. Gottstein and Kayser specify an epidemiological principle, the necessity to select samples from well-defined populations. They are concerned to test people in their target group who are drawn from the same workplace, who have worked there uninterruptedly, and who therefore have a common occupational environmental background. They are concerned to exclude those with auditory impairments attributable to factors other than noise at work.

Gottstein and Kayser are careful to report results in ten-year ranges, so as to show the comparability of the two samples on that vital dimension, time.

The principal and striking finding they point to is the increasing impairment sustained by the metal workers through the course of their working lives.

Gottstein and Kayser attribute to noise the reduction in the metal workers' hearing to occupational noise on the following grounds. The workers themselves attribute the gradual worsening of their hearing to the noise at work; there is an increasing impairment in bone conduction as well as air conduction of sound (as tested with a Poliit; with a coupler). Wax in the ear canals does not relieve the symptoms to any material extent, and in any case blockage of the ear canals by wax was found in only 10 of the men; there are reports of tinnitus from a moderately high proportion (24 per cent) of them; physical abnormalities detected in the ears of the majority in the "bad" and "fairly bad" categories, and in others, were not considered causal because several men in these categories had no such abnormal signs.

Gottstein and Kayser noted that workers in these and similar noisy trades are exposed to many other agents that can affect hearing, such as draughts, rapid temperature changes, and dust, but concluded that these factors must be seen as subordinating because bricklayers are equally exposed to them. The probable cause of the deafness in their view is a "dulling of the perception of the nerves of hearing (p. 206)."

Thomas Barr's "Inquiry": Epidemiology Extended and the Use of Self-Report

Barr's main method of testing was his pocket watch. He reports this 'being heard when the hearing is normal 38 inches from the ear, the distance for normal hearing was arrived at. I Barr does not say, but we can fairly confidently take it that he relied on his own hearing capacity to determine his yardstick, literally and metaphorically, of 'normality'.

Barr used his watch first to test the hearing of 100 men working as boilermakers in two Glasgow shipyards. He gave data on their age and occupational experience.

Barr's next step is in advance in one way on Gottstein and Kayser's epidemiology in that he studied two further occupational groups, comparable in age but engaged in less acoustically traumatic trades. These two comparison groups were 100 iron-workers (working in a moderately noisy environment) and 100 letter-carriers (working in the environment at large). He tallied the number of inches distance from each ear that each man could detect the watch-tick. He then expressed the total for each occupational group as a proportion of 7,200 inches, the total value of distances 100 men with normal hearing should hear the watch-tick opposite each ear.

Apart from the watch, Barr used whispered and loud speech. He appraised the capacity to hear in a public place by asking the men to recount for themselves their problems.

Barr takes up the cause of those with hearing impairment, estimating their incidence at 10 per cent of the adult population. He pleads for better acoustical design of public meeting places, closer concern by clergy and the like over the clarity of delivery of their message, not the least aspect of which would be to refrain from growing bushy beards and moustaches.

For diagnosis and clinical description, Barr relies on the observation of bone conduction deficit to conclude, in common with Gottstein and Kayser, that the cause of the dysfunction is overstimulation of the auditory nerve. But he also reports on the differential sensitivity for tones of different wavelength causing high frequency loss, due in his judgement to the basal nerve cells being the more damaged as they are acted upon by the shrill high-pitched sound of boilermaking. He records that the incidence of permanent tinnitus is much less than in a typical clinical sample.

Barr describes the course of the deafness, noting the wide inter-individual variability in rate of onset, and draws the typical picture as one of a fairly immediate experience of reduction in hearing power, for instance, among apprentices. He appreciates the temporary overlay (now referred to as temporary threshold shift, TTS) manifested by the occurrence of rapid recovery in hearing, through the course of an evening, in the early days of exposure, by recovery over the holiday period even in men of more advanced occupational exposure, and by recovery due to longer-term 'time-out' from the trade. He devotes considerable and careful
attention to personal protective devices as a means of reducing the risk of damage, advocating indiarubber plugs and cotton plugs smeared with vaseline. He warns about problems of fitting, of irritation of the meatus, and of motivation to use such devices among industrial workers.

Occupation-based vs Dose-response Epidemiology

The dose-response method for noise is represented best in the work of Burns and Robinson (1970), which sought to observe the course of hearing injury through time, as measured by the present-day equivalent of Barr’s watch-tick and Politzer’s acoumeter, pure-tone audiometry. The prospective measurements of Burns and Robinson revealed no consistent and characteristic pattern of decline in hearing in individuals, though a pattern could be discerned in pooled data. They ruled out a prospective approach in favour of retrospective, that is, correlation between measured injury and occupational history. Retrospection was the approach used by Gottstein and Kayser and by Barr. But Burns and Robinson’s retrospection went further: it deduced individual noise exposures from occupational histories by extrapolating noise data from field studies. This is the procedure of dose-response epidemiology, which transforms occupational histories into doses of noise. Doses of noise are then used to compute dose-response relation. Dose-response relation is then used to predict risk.

It can reasonably be claimed that the concrete virtue of this abstract, multi-step approach to assessment of risk to hearing from noise is in allowing any occupational environment to be assessed so as to predict likely risk to the hearing of those at work within it. But the virtue has diminished because there have grown up two such predictive schemes. While occupation-based epidemiology could not assist researchers in clearing up the uncertainty of two rules, adequate understanding might have prevented it from developing in the first place.

Understanding lies in recognition that dose-response relation introduces two more major uncertainties additional to those inseparable from occupation-based epidemiology. One uncertainty arises in the assumed correlations between noise measurements and occupational history in the original research; the other, in the extrapolation to individual exposure from dose-response relation in the application of the research findings.

The virtue of occupation-based epidemiology is in allowing particular trades within larger occupational groups to be identified as undoubtedly associated with risk to hearing.

Self-report versus Test-based Assessment of Hearing Handicap

Just as dose-response relation relies on multi-step abstractions, so do many schemes of assessment of hearing handicap. “Hearing handicap” has come to be defined typically as “any disadvantage in daily life arising out of hearing impairment”, and is assessed for purposes of rehabilitation and compensation. Barr’s notion of “interference with social comfort or usefulness” is consistent with well-accepted and current definitions of hearing handicap. Whereas Barr relied on his respondents’ self-appraisal of their hearing in a typical social setting (the public meeting) to establish a measure of social comfort, the approach which has developed since his time has relied more and more on performance in tests, such as pure-tone audiometry, of the type Barr used merely to measure loss of hearing power (watch-tick, whispered and loud speech). The reasoning is that such tests have the virtue of being standardized instruments so that any one person’s performance is directly comparable with that of any other.

A major problem facing this supposedly scientific approach has been to determine what a given level of reduced performance means for “disadvantage in daily life”. Results from tests have to be correlated with every-day understandings; there is a translation problem attending the use of tests.

The presumed disadvantage of self-appraisal has been its subjectivity. How can one rely on what individuals report about their subjective experience, and how can personal and private meanings not intrude upon the perceptions individuals describe? Performance in a standard test presents the appearance of avoiding the disadvantage of subjectivity. But subjectivity is not circumvented by use of standardized tests, because the translation problem is not circumvented. Rather, the moment of subjective appraisal is shifted from the occasion of testing to the occasion of that translation. Somehow, someone has to locate the performance measure on a scale of handicap. Thanks to standardized tests, it has transpired that the person doing that in the examiner, not the person undertaking the test. The subjectivity of each testee is replaced by another subjectivity which establishes or refers to an arbitrarily calibrated scale of handicap and locates each person’s performance on it. The situation is analogous to Barr’s use of his own hearing as the standard of normal for all listeners. The situation is also analogous to the use of dose-response relation with its several steps of assumption.

References


APPLICATION OF CONTEMPORARY EPIDEMIOLOGIC METHODS TO INDUSTRIAL HEARING CONSERVATION

John Erdreich and Linda S. Erdreich

National Institute for Occupational Safety and Health, 4876 Columbia Pkwy., Cincinnati, OH 45226 USA

One hundred years ago Thomas Barr published the first report that examined the risk factors that relate occupational noise exposure and the development of noise-induced hearing loss. The celebration of this seminal paper in occupational health should serve to focus on the direction future efforts in the field of occupational acoustics should take. This is an opportunity to ask, What remains to be done in research on noise-induced hearing loss?

The purpose of this paper is to provide an answer to that question. Simply providing evidence of a causal relation between the agent and effect represents the beginning of the process to understand and eliminate the adverse effects of any hazardous exposure. As an example, the conclusions developed by Barr regarding the length of time men in different jobs were free of impairment could not be generalized to industries other than bollermaking because of the lack of quantitative exposure data. Early studies of noise exposure and hearing loss were generally concerned with determining the degree of impairment after a defined level and duration of exposure. These cross-sectional studies have provided the basis on which we can project the percentage of a population at risk from specific exposures. However, with the mandate for audiometric testing in the United States, there is a need for studies evolving that can provide further insights to the protection of hearing in industry.

TECHNIQUES

Examination of these data reveal several factors that render traditional analytical techniques useless. First, many individuals are lost to follow-up, that is, they do not all have the same number of tests. For example, in the data presented here, 2651 people had one test, 2368 had five tests, and only 837 had ten tests. Second, each person in the database is not followed to the point at which a Hearing Loss or a Standard Threshold Shift develops as would be appropriate for a standard epidemiologic cohort analysis. The data is termed "censored" which means that participants may start at different times and may reach the outcome at different times, if at all. In this case only 446 of 2651 participants developed a hearing loss by the end of the study. We would like to examine the effect of this exposure on males and females and the effect of the age at which the first exposure occurs on the development of hearing loss. Linear regression techniques are not suitable for analysis of information that contains dichotomous or categorical outcomes. Survival analysis is a technique which will both treat censored data and help us examine the pattern of events over time. Temporal trends in developing hearing loss may aid our understanding of the mechanism of noise-induced hearing loss and may suggest strategies for its prevention. Erdreich and Erdreich (1983) suggested that there was much to be learned from applying survival analysis and life table techniques (Lee, 1980) to audiometric data. The intent of the analysis is to plot survival curves and to produce hazard functions for various noise-exposure groups, examine the probability that after being exposed for some test interval they will complete the next interval (in this case one year) without suffering the outcome. Effects of sex and the analytic variables can be evaluated with these techniques. We start by creating outcome variables and performing descriptive analyses in order to select appropriate survival methods.

ANALYSIS OF AN INDUSTRIAL DATABASE

To examine the temporal aspects of industrial hearing loss we used data from approximately 2650 individuals for whom we have as many as fourteen sequential audiometric tests. Age, sex, and noise exposure levels are among the data included for each of the records.

Two outcomes were chosen for examination: the first is a hearing loss averaging 25 decibels at 1KHz, 2KHz, and 3KHz in either ear. The second outcome is the identification of a Standard Threshold Shift (STS) as defined by any of the three rules proposed in the 1981 version of the Hearing Conservation Amendment to the United States Occupational Safety and Health Administration Noise Regulations. The definition of STS is: (1) for employees whose hearing is within 25 dB of audiometric zero and who have not suffered an STS a 20dB shift at any frequency. (2) For employees whose threshold is 22dB at any test frequency, the criterion is 10dB or greater at 1000Hz and 2000Hz, 15dB at 3000Hz and 4000Hz and, and 20dB at 8000Hz. (3) For employees with a loss worse than 25dB average at 1000, 2000, and 3000Hz, the criterion is 10dB at any test frequency. Because it allows less loss for workers with identified threshold shifts, an STS identified by each rule was considered to be different only with others identified by the same rule.

Four hundred and forty-six (446) people were identified who had a hearing loss according to the definition. One thousand three hundred and one person (1318) people were identified who had a Standard Threshold Shift by either of the three computational rules. Because we were interested in examining temporal trends in hearing loss, we identified the number of the test on which each person was first identified as having a hearing loss and the number of the test on which each person identified as having an STS. The largest percentage of cases of hearing loss are identified on the test
one. This reflects the fact that the hearing conservation program from which these audiological data are derived was implemented after many of the participants had been hired.

The high percentage of STS identifications occurring on the second test is of more interest. Learning on sequential tests has been postulated to account for a small improvement in thresholds, but this trend is in the opposite direction. However, there are different numbers of individuals taking each test so it is necessary to calculate the percentage of those taking a given test who suffer a hearing loss or an STS.

Adjusted rates of STS and hearing loss identification were calculated based on the number of subjects in each test. These results are shown in figures 1 and 2. Adjustment for the number of tests in each category reduces the differences between the percent of cases of the identified outcome in later tests (three through ten) but the pattern described for the early tests in the unadjusted data remains.

IMPLICATIONS FOR INDUSTRIAL HEARING CONSERVATION

This initial analysis of the data from a small sample of people in a hearing conservation program illustrates that there is still much to be gained beyond the demonstration of Dr. Barr that noise causes hearing loss. The finding that the majority of early impaired thresholds occurs in the first year of exposure is evidence that effective techniques to preserve hearing must be implemented very early in a program. Furthermore, if these trends are supported in other industries and other programs, we may conclude that in the early years of exposure it is of benefit to both the worker and to his employer to test hearing semi-annually to assure early identification of developing problems.

REFERENCES


PROCEDURES FOR THE RISK APPRAISAL FOR NOISE-INDUCED HEARING LOSS AT THE LEVEL OF A SINGLE INDUSTRIAL SETTING

R. Héroux, R. Phaneuf

GAMU, Université de Montréal, Pavillon Marguerite d’Youville, C.P. 6128, Succursale A, Montréal, Québec, Canada, H3C 3J7

Health professionals responsible for preventing occupational hearing loss in industry need scientific criteria to operationally define priorities for noise exposure reduction. This can be achieved by assessing the risk of hearing loss attributable to the exposure in the population. One approach of risk appraisal is the measure of exposure levels: it allows to order the severity of the exposures for different jobsites, job categories or departments. Estimates of the long term representative daily noise dose can be used to predict hearing loss as proposed in the ISO-DIS/1999 (1). But there are serious limitations to this procedure, both in terms of its internal and external validity. These limits are discussed below and a complementary approach based on an epidemiological procedure is proposed.

1. LIMITS TO THE INTERNAL VALIDITY OF THE DOSE-RESPONSE RELATIONSHIP

The proposed standard is based on the results of cross-sectional studies on noise-induced hearing loss. The evaluation of the exposure relied on an assumption: the workers selected had been exposed to the one and only noise condition that was appraised at the time of noise measurement. Errors of over- and under-estimation of the exposures may have occurred non-intentionally. The effect of this non-differential misclassification bias is to underestimate the true measure of association. (2)

The second threat to the validity of the standardized dose-response relationship is the possibility of a selection bias. In the original studies on which it is based, no attempt was made to examine whether the people recruited to represent the higher exposure groups might have evolved from a natural selection procedure. We can consider that at least some of the most susceptible individuals to noise-induced hearing loss would have removed themselves from exposure to the noisier jobs. The individuals involved in these studies are presumably of greater resistance to the hearing loss on an average. Therefore, the selection effect is to reduce the true association between exposure and health impairment. (3)

Thus, effects of misclassification and selection biases presumably reduced the real measure of association, particularly for the higher exposures.

2. LIMITS TO THE EXTERNAL VALIDITY OF THE DOSE-RESPONSE RELATIONSHIP

The data on which the ISO-DIS/1999 is based were obtained from studies of the effects of exposures essentially characterized by steady-state broadband noise during regular 8-hour days. Risk estimation cannot be generalized to exposures that seriously depart from the original conditions, without some degree of uncertainty. Thus, the model assumes a yearly exposure of 2000 hours; but it is not unusual for many industrial workers to accept an additional 400 to 800 hours of work. Several workers also are on a 12-hour schedule, a condition that has not been studied yet. There remains several outstanding questions relating to the effects of impulse noise, very low frequency noise, tonal noise and exposures to high sound levels ($L_a > 100$ dB) during only small portions of the workday. The latter situation is common in industrial settings where reduction of the exposure relies on the use of sound treated control rooms.

The normalized dose-response relationship also excludes possible interactions between the effects of noise and other physical, chemical or biological conditions, such as exposure to vibrations, non-noise occupational irritants, which could contribute to the hearing loss. (4)

Moreover, there are difficulties in determining the noise exposure for everyone in a plant because of work procedures. For production workers, it is usually strongly varied. However, for significant portions of the work force, in the maintenance department, reliable estimates of exposures may be very difficult (5); job assignments can involve almost any site in the plant for periods of time that are determined by the specific needs for trouble-shooting and repairs.

This situation raises another type of uncertainty in the risk appraisal, namely the integration of highly variable daily, monthly, or yearly exposures into a long term average exposure index. Validation of the equal energy hypothesis has not been demonstrated for these extreme cases. Finally, related to this question, serious doubts can be raised on the validity and realism of long term average exposures in view of the increasing internal mobility in industry. It is now an exception for workers to hold a job at a given site for 10, 15 or 20 consecutive years. For example, in a foundry employing between 100 and 400 workers, an average of 10 individuals were replaced yearly; over a 10 year period, it implied 100 new workers and nearly 1000 job shiftings (6). This "domino" effect is governed by rules adopted in the collective agreement; changing from one job to another is motivated by a proportion of at least 90% by higher wage opportunity and day shift assignment. Consequently, accurate noise dose measurement at a given jobsite cannot provide a valid estimate of the long term average exposure of an individual, unless very complex analyses of the job shifting within the plant are performed. In other words, the traditional risk assessment as proposed in the ISO-DIS/1999 relies on practice on an unlikely postulate: namely that a given worker will occupy his current job for the next 10 or even 20 years. The problem of high internal mobility is certainly not restricted to a small number of plants and industrial sectors and it will likely become a standard with the implementation of new technologies.

Therefore, the original limitations of the inferences about noise-induced hearing loss is seriously restricted when applied to the actual situation of an industrial setting. Because of these limitations, we wish to consider to suggest an epidemiological approach as a complementary procedure of risk assessment, which can contribute to give rigorous foundations to the definition of noise reduction priorities.

3. A COMPLEMENTARY APPROACH TO RISK EVALUATION

A second possibility of risk appraisal is to undertake a cross-sectional study of non-disease-induced hearing loss with the current exposure or job classification. The ratio of prevalences of hearing losses can be reviewed as a function of relevant categories of exposures, such as the departments or the trades within the plant.

Provided that the turnover is low and that the mobility is limited between departments, workshops or trades within the plant, it allows to take into account the actual exposure conditions that prevailed in the particular firm considered. It should be particularly useful for assessing the risk of the exposure conditions that seriously departs from those considered in the original epidemiological studies included into the ISO-DIS/1999 data base.
A procedure has been proposed to obtain a reliable estimate of the prevalence of significant hearing losses (SII) in this context. (7) SII is defined as a loss at a given audiometric frequency that is unlikely due to the effect of age (i.e. above the 90th centile, ref: ISO-1029).

The prevalence odds-ratio (POR) can be computed as the measure of association between hearing loss and the exposure index. It is calculated on results of those workers whose hearing status is not indicative of a pathology not related to the working environment. A chi-square statistic can be used to test whether the observed hearing losses occurred merely by chance. The confidence limits of the POR can be computed by a test-based method. (3) This is illustrated in Table 1.

Table 1. Prevalence odds-ratio (POR) and chi-square statistic (X^2) applied to measures of significant hearing loss (SII) in a noisy plant.

<table>
<thead>
<tr>
<th>SII</th>
<th>Normal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>noisy dept.</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>reference dept.</td>
<td>c</td>
<td>d</td>
</tr>
<tr>
<td>N1</td>
<td>N2</td>
<td>N</td>
</tr>
</tbody>
</table>

\[ \chi^2 = \frac{(ad - bc)^2}{n1 \cdot n2 \cdot n1 \cdot m2} \]

95% confidence limit of POR = POR \pm (1.96 \times \chi^2)

An example of the above procedure is given in Table 2 with data from a steel mill. The reference group comprised the employees from the administrative and quality control departments of the plant.

Table 2. Prevalence odds-ratio (POR) of significant hearing loss for three departments of a steel mill.

<table>
<thead>
<tr>
<th>Department</th>
<th>SII</th>
<th>Normal</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>reference dept.</td>
<td>2</td>
<td>24</td>
<td>26</td>
</tr>
<tr>
<td>finishing</td>
<td>14</td>
<td>66</td>
<td>60</td>
</tr>
<tr>
<td>foundry</td>
<td>8</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td>maintenance</td>
<td>32</td>
<td>33</td>
<td>65</td>
</tr>
</tbody>
</table>

\[ \chi^2 = \chi \cdot \text{Probability of Confidence Interval} \]

The results from Table 2 show that the overall risk of hearing loss in plant is significant; the association is evident for the maintenance department, with the highest POR. The second most important department in terms of risk is the finishing department.

There are limitations with this procedure as in any cross-sectional study. The size of the population will determine the precision of the risk estimate. It is subject to selection biases that could affect differentially the various department or trades within a plant. Mobility across departments could lead to underestimate the risk. This is especially true in plants where the most demanding jobs, which frequently involved the most severe exposure, are systematically occupied by the youngest workers; a certain number of years of experience makes them eligible to a change of assignment for a job category that may be less at risk. This results in a misclassification bias that reduces the prevalence odds-ratio of the overall classification. Finally, the exposure time may be critical to the risk estimation; stratification of the population according to the length of service in a department or trade would give a better measure of association, provided that the number of workers is sufficiently large.

An adjusted odds-ratio can be computed for a comparative analysis.

This approach can also be applied to data from different plants in a given industrial sector, across trades in a group of plants and across industrial sectors, provided that there is a valid reference population.

4. CONCLUSION

In plants where the exposure is easily quantifiable and where there is little turnover of the personnel, the assessment of noise-induced hearing loss is simple and straightforward. The procedure outlined in the proposed draft of international standard, based on environmental assessment, should suffice; though it could underestimate the real hearing loss, the priorities for noise reduction would not be confused.

But in most plants, both environmental and health assessments are essential to obtain valid indicators of the risk.

REFERENCES

PERTES D'AUDITION PAR EXPOSITION A DES BRUITS INDUSTRIELS DE L'Aeq COMPRIS ENTRE 85 ET 90 dB(A)

1. Thierry, C. Meyer-Bisch, H. Arbey

Institut National de Recherche et de Sécurité
B.P. 77, 54501 Vandœuvre, France

Une enquête épidermologique transversale a été réalisée dans un grand atelier de tôlerie automobile pour quantifier l'effet sur l'audition d'une exposition professionnelle à des bruits de niveaux acoustiques continus équivalents compris entre 85 et 90 dB(A).

1 - Méthode

Les 1200 travailleurs de l'atelier ont été interrogés par questionnaire sur leurs histoires professionnelles, médicales et extra professionnelles. Parmi eux, on a sélectionné 234 personnes réunissant les caractères suivants :
- sexe masculin,
- pas d'exposition professionnelle, autre qu'en tôlerie, à des bruits proches ou supérieurs à 90 dB(A),
- pas de pathologie ni d'antécédent auditif,
- pas d'utilisation de protecteurs auditifs.

Cet atelier, créé il y a 20 ans, n'a pas été, entre temps, l'objet de modifications technologiques majeures.

Une étude sonométrique détaillée a été réalisée dans l'atelier. Les mesures ont été effectuées en 164 points répartis dans l'atelier, représentant la totalité des postes de travail, à l'aide d'un enregistreur magnetique et d'un sonomètre intégrateur de précision. Les bruits de cet atelier sont en partie impulsionnel et fluctuants. Le L_Aeq moyen pour l'ensemble de l'atelier est de 87,8 dB(A) (α = 3,2 dB(A)). Les L_Aeq moyens des divers postes de travail ne diffèrent jamais de plus de 5 dB(A).

Les examens audiométriques manuels ont été réalisés avant le début du poste de travail, dans des conditions normalisées (étalonnage conforme à la norme ISO R 389 ; absence de TTE ; population sélectionnée selon la méthode ISO/DIS 1999).

Quatre classes homogènes pour l'âge et la durée de l'exposition ont été définies. Leurs effectifs figurent dans le tableau 1. Elles regroupent 183 travailleurs sur les 234 sélectionnés.

2 - Résultats

Le tableau 1 représente la distribution statistique des pertes auditives des travailleurs des quatre classes, moyennées sur les fréquences 3, 4 et 6 kHz (indice "ML 346").

<table>
<thead>
<tr>
<th>HL 346 (dB)</th>
<th>Classe [Age/Durée d'exposition / années)</th>
<th>Effectif</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[29/7]</td>
<td>[32/12]</td>
</tr>
<tr>
<td>0.90</td>
<td>41</td>
<td>43</td>
</tr>
<tr>
<td>0.50</td>
<td>7.8</td>
<td>9.4</td>
</tr>
<tr>
<td>0.10</td>
<td>16.5</td>
<td>20.1</td>
</tr>
<tr>
<td>Moyenne</td>
<td>31.6</td>
<td>42.9</td>
</tr>
<tr>
<td>σ</td>
<td>18.1*</td>
<td>23.6</td>
</tr>
</tbody>
</table>

Tableau 1 : Distribution statistique des pertes auditives "ML 346". (* différence significative entre cette valeur et les 3 suivantes : test T, p < 0.05).

Les résultats de cette étude sont comparés à d'autres données déjà publiées par l'INRS en 1980 [1]. La figure 1 présente les valeurs médianes des pertes auditives pour les 2 classes extrêmes de l'atelier, [29/7] et [40/18]. Elles sont comparées aux pertes médianes de populations témoins d'âges moyens 30 et 40 ans, puis à celles de populations exposées à 95 dB(A) pendant 10 et 20 ans.

Figure 1 : Valeurs médianes (30 %) des pertes d'audition pour 2 classes de l'échantillon de tôlerie, comparées à des valeurs obtenues par l'INRS en 1980.

La figure 2 présente les résultats observés pour la classe [40/18] selon l'indicateur HL 346. Ces résultats sont comparés aux estimations des pertes auditives fournies par le projet de norme ISO/DIS 1999 (1982), pour une population masculine bien sélectionnée (pas de pathologie), d'âge 40 ans et exposée 20 ans à des niveaux sonores L_Aeq de 90, 95 et 100 dB(A).

Figure 2 : Distribution statistique des pertes d'audition (HL 346). Résultats obtenus dans l'atelier pour la classe [40/18] comparés aux estimations fournies par la norme ISO/DIS 1999 pour la classe [40/20] et pour des niveaux sonores de 90(•), 95 (♠) et 100 dB(A) (▲).
J - Discussion

Ces résultats confirment l'existence d'un risque pour l'audition dû à l'exposition à des bruits de niveaux moyens de 8a eq compris entre 85 et 90 dB(A). Ce risque se traduit par l'aggravation observée sur l'indicateur précoce, "HL 346", quand la durée d'exposition aux bruits s'accroît de 7 à 12 ans. Cette aggravation est de 11,3 dB pour les 10 % de la population les plus atteints. Si on considère les écarts entre les valeurs moyennes de cet indicateur, le test de Student ne permet pas de faire apparaître d'écarts significatifs entre les classes [32/12], [36/15], [40/18]. Toutefois, il existe une différence significative (p < 0.05) entre la classe [29/7] et les trois autres classes.

Si on ajuste sur l'âge ces résultats en utilisant la table préconisée par l'OSHA en 1981 [2], les corrections obtenues sur HL 346 sont respectivement de 9, 10, 12 et 14 dB pour chacune des classes considérées. Ainsi, entre 29 et 32 ans, l'effet de l'âge correspondrait à une aggravation de 1 dB qu'il faudrait retrancher aux valeurs observées pour obtenir les valeurs corrigées. Cette aggravation "normalisée" atteindrait ainsi 10,3 dB sur les fréquences sensibles 3, 4, 6 kHz, pour les 10 % de la population les plus atteints. L'analyse actuelle de ces résultats ne permet pas de discuter la pertinence d'un critère d'atteinte auditive [3]. Il est important, à ce sujet, de prendre en compte la diversité des modes d'ajustement des pertes auditives sur l'âge, dont le choix influe sur les résultats.

Les valeurs médianes des pertes d'audition (fig. 1) font apparaître chez les travailleurs de tôlerie que le maximum de la perte auditive se situe à 6 kHz et non à 4 kHz comme on le constate dans la population exposée à des bruits de 95 dB(A). Enfin, la comparaison avec les estimations de la norme ISO/DIS 1999 montre (fig. 2) que pour l'indicateur HL 346, les pertes auditives observées dans cet atelier sont confondues avec celles du niveau sonore 95 dB(A).

Si cet écart ne peut pas être expliqué par un mauvais étalonnage de l'audiomètre, il pourrait provenir du caractère impulsionnel des bruits de tôlerie (chocs entre tôles et parfois martelage). Ceci pourrait confirmer l'hypothèse selon laquelle le caractère impulsionnel des bruits est pénalisant, ainsi que des travaux récents effectués dans le secteur de l'estampage semblent l'indiquer [4], [5]. Il n'est toutefois pas possible d'aller plus loin dans cette analyse de nos résultats car le protocole épidémiologique de l'étude n'avait pas été prévu pour tester cette hypothèse.

Nouss remercions S.B. Luong et M. Steunou pour leur participation active à ce travail.

Bibliographie


AN ANALYSIS OF THE RATE OF GROWTH OF HEARING LOSS IN THE CANADIAN FORCES

S.E. Forshaw

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
Downsview, Ontario, M3M 3B9, Canada.

INTRODUCTION

Hearing threshold levels (HTL) are classified in the Canadian Forces (CF) for the purposes of enlistment, career assignment, and computer medical records by the four hearing categories shown in Table 1. The analysis reported in this paper was part of a study carried out using these categories to assess the effectiveness of the CF hearing conservation program across military occupations (1).

TABLE 1

CANADIAN FORCES HEARING STANDARDS

CATEGORY HEARING STANDARD
H1 HTL not greater than 30 dB between 500 and 6000 Hz in both ears.
H2 HTL not greater than 30 dB between 500 and 3000 Hz in both ears.
H3 HTL not greater than 30 dB between 500 and 2000 Hz in the better ear.
H4 HTL not greater than 50 dB between 500 and 2000 Hz in the better ear.

PROCEDURE AND RESULTS

Linear regression equations were computed for each Military Occupation Code (MOC), treating the percentage of personnel in hearing category H1 within five-year age groups as a random variate. Due to the relatively small numbers in certain of the MOCs, the data from these MOCs were combined with data from at least one other MOC with the same regression-line slope or from a related trade or career. Linear regression equations were then computed for the grouped MOCs. From these data, the relative risk-to-hearing for each MOC or grouped MOCs were ranked in order of regression-line slope.

To estimate the intensity of the noise experienced on a daily basis that would produce the rate of growth of hearing loss observed, the MOCs were grouped into populations I and II representing occupations with the smallest and the largest percentage decreases in hearing category H1 respectively, as represented in Figure 1.

Group I represented the upper 6th percentile of the CF population with respect to regression-line slope (11.6 per cent decrease per decade based on the median regression-line slope of the MOCs in the Group), and included occupations such as medical officers, nurses, fire control technicians, terminal equipment and radio technicians. Group II represented the lower 13th percentile of the CF population with respect to regression-line slope (25.0 per cent decrease per decade, based on the median regression-line slope of the MOCs in the Group), and included mainly infantrymen and air and land weapons technicians.

The age distribution of the personnel within each Group is shown in Figure 2. Analysis of the distributions indicate a significant variation between groups (p<.01), the deviations from expected values for uniform distribution being greatest below age 35. This reflects the fact that a high proportion of persons join the infantry in their late teens and leave the service in their mid twenties (the 50th percentile age is about 23), whereas persons in the medical services enter the CF in their early twenties and remain in the service for a longer period (the 50th percentile age is about 31).

Within the Group I and II populations, the 50th and 90th percentile ages of persons in hearing category H1 were estimated using the median regression-line slopes. The results indicate that at age 23 to 24, 90 per cent of the persons in the Groups were in hearing category H1 (see Table 2). By age 30, this had dropped to 80 per cent in Group I. The relatively small number of persons in the CF above age 50 precluded a meaningful estimate of the 10th percentile age in either Group, and the 50th percentile age in Group I. At age 50, about 90 per cent in Group I remained in hearing category H1.
TABLE 2
ESTIMATED 50TH AND 90TH PERCENTILE AGES
FOR PERSONS IN HEARING CATEGORY H1, BY MEDIAN
REGRESSION-LINE SLOPE IN GROUP POPULATIONS I AND II

<table>
<thead>
<tr>
<th>POPULATION</th>
<th>50th</th>
<th>90th</th>
</tr>
</thead>
<tbody>
<tr>
<td>GROUP</td>
<td>PERCENTILE</td>
<td>PERCENTILE</td>
</tr>
<tr>
<td>I</td>
<td>&gt;50 years</td>
<td>24 years</td>
</tr>
<tr>
<td>II</td>
<td>39 years</td>
<td>23 years</td>
</tr>
</tbody>
</table>

To compare these estimates with published epidemiological data, the relationship given in ISO Standard 1999 (2) between permanent threshold shift (PTS), noise exposure and presbycusis was used to estimate the 50th percentile ages of otologically-normal males who would exceed the criteria given in Table 1 for hearing category H1 when exposed to equivalent noise levels of 85, 90 and 95 dBA (see Table 3).

TABLE 3
ESTIMATED 50TH PERCENTILE AGE AT WHICH THE HTL AT
4000 AND 6000 HZ EXCEEDS THE CRITERIA FOR CF HEARING
CATEGORY H1 ASSUMING NOISE EXPOSURE BEGINS AT AGE 18

<table>
<thead>
<tr>
<th>EQUIVALENT NOISE LEVEL</th>
<th>50th PERCENTILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>85 dBA</td>
<td>57 years</td>
</tr>
<tr>
<td>90 dBA</td>
<td>51 years</td>
</tr>
<tr>
<td>95 dBA</td>
<td>40 years</td>
</tr>
</tbody>
</table>

Comparison of the estimated 50th percentile ages given in Tables 2 and 3 suggests that the rate of growth of hearing loss observed generally among persons in the least hazardous noise-exposure occupations could be the result of daily exposure to noise with an equivalent level not greater than 85 dBA, and in the most hazardous noise-exposure occupations, from daily exposure to noise with an equivalent level of about 95 dBA.

DISCUSSION

Persons exposed to noise on a regular basis can develop varying degrees of hearing loss. With the exception of exposure to blast, high intensity impulse noise, and extremely intense steady-state noise, the growth of permanent hearing loss progresses over months, years, or decades of exposure. Of course, military noise exposure does not occur at a constant rate. Short-term variations depend on day-to-day and week-to-week training and operational activities. Long-term variations depend on the noise exposures associated with career postings (e.g., squadron or regiment versus staff postings). The foregoing exposure estimates may be viewed, therefore, as equivalent unprotected-exposure noise levels that, if sustained on a regular basis, would produce the observed hearing losses that have occurred from variable exposure.

Certain simplifying assumptions have been made in this study. It is assumed, for example, that the non-occupational noise exposure experienced by non-military personnel (e.g., due to recreational activity or non-military employment) is negligible compared to their military noise exposure. This is questionable for persons in the Group I MDCs, so that their rate of growth of hearing loss due to military noise exposure may be underestimated.

It is assumed also that any longitudinal heterogeneities in the noise susceptibility of the CF population due to differences in sex distribution, and/or deviations from the linear-regression model assumed for the CF data, have not had major effects on the estimates. To the extent that these assumptions are not true (e.g., if a significant number of overly susceptible individuals leave high-noise-exposure occupations because of the rapid onset of hearing impairment), the rate of growth of hearing loss in Group II may be underestimated.

The use of ISO Standard 1999 for noise with instantaneous sound pressures in excess of 140 dBA (e.g., weapons) involves considerable extrapolation. Any assumed equivalency in terms of energy of impulse and steady-state noise may not be entirely valid. Price has suggested, for example, that there is a critical intensity level for impulse noise above which noise-induced threshold shifts occur more rapidly (3).

Suffice it to say, the foregoing noise-exposure estimates and rates of growth of hearing loss should be viewed simply as statistically derived population trends, and should not be used to predict or assess the hearing impairment of individual persons.

At the same time, the results for Group II are supported by previous findings: high-frequency hearing loss is prevalent among military personnel in general, and infantry, artillery, and armoured-corps troops in particular. Two obvious explanations are available. One is the common failure of personnel to use hearing protection consistently during weapon firing; the other is the very intense and continuous noise exposure experienced by these troops during training and operational exercises.

Undoubtedly, the classification of population hearing data as gross a descriptor as military hearing category (as opposed to the HTL at each audiometric test frequency, masks all but very large changes in population hearing level trends. Moreover, the procedure of maintaining an audiometric data base solely by hearing category precludes comparison of such data with the many data bases in which growth of hearing loss is available directly in HTL.

CONCLUSION

This study was undertaken to determine the rate of growth of hearing loss in the various CF occupational classifications. Comparison of the CF HTL data with the epidemiological data published in ISO 1999 suggests that the PTS observed in the least hazardous noise-exposure occupations could be the result of equivalent daily exposures not greater than 85 dBA, and in the most hazardous noise-exposure occupations, from equivalent daily exposures of about 95 dBA.

REFERENCES

L'INFLUENCE DE L'ORIENTATION SCOLAIRE SUR L'ACUITÉ AUDITIVE DES ÉLÈVES (1)

L. Paré et F. Filion

DSC De Lanuaille, 1000 Boul. Ste-Anne, Joliette, Québec, J6E 6J2

INTRODUCTION

Peu d'études ont porté sur les effets de l'exposition au bruit sur l'acuité auditive des élèves inscrits en apprentissage professionnel. Woodford et O'Farrell (2) ont comparé des sujets des deux sexes fréquentant ou non des ateliers bruyants. Ils déclèrent une perte auditive (seuil > 15 dBHL) chez 61,6% des sujets. Les élèves exposés aux ateliers bruyants montraient un pourcentage plus élevé d'atteintes.

Comme il n'existe pas de données canadiennes de référence, notre objectif premier est la description des seuils auditifs de l'ensemble de notre population. Notre second objectif est de voir s'il y a une association entre l'exposition au bruit en cours d'apprentissage et l'acuité auditive des élèves. En dernier lieu, nous tenterons de distinguer quel est le meilleur traitement des données d'exposition pour prédire l'acuité auditive. Au delà de la dose d'exposition cumulée (Eₐ), le mode temporel de cumulation de l'énergie pourrait être une meilleure façon de tenir compte de l'exposition dans l'évaluation du risque d'atteinte.

MÉTHODOLOGIE

1. Approche expérimentale

Nous avons choisi de mesurer et de comparer l'audition de deux groupes d'élèves: un d'orientation professionnelle (exposé au bruit) et l'autre d'orientation académique (non exposé au bruit). La similitude des sujets quant aux autres facteurs influençant l'audition n'est connue qu'à posteriori. Le contrôle de ces facteurs est effectué au niveau de l'analyse statistique.

2. Population


3. Procédure

Le protocole d'examen consiste en un questionnaire (entretien), un examen tympanométrique et un examen audiométrique (0,5 à 6 kHz). Le questionnaire identifie le profil scolaire, les expériences de travail, le service militaire, les activités bruyantes de loisirs et les antécédents médicaux (indice d'antécédent personnel). L'exposition annuelle (Eₐq) et l'exposition cumulée en cours d'apprentissage scolaire (Eₐ) sont établies à partir de relevés sonométriques en atelier et d'informations des enseignants sur le temps d'exposition.

RÉSULTATS

1. Seuils auditifs des 16-20 ans

Le tableau 1 présente les seuils auditifs (moyenne des deux oreilles) de l'ensemble des sujets. Les seuils moyens de la présente étude sont légèrement supérieurs, à toutes les fréquences, aux seuils obtenus dans l'étude de Roche et al. (3). L'écart-type de cette dernière étude est cependant plus élevé.

2. Exposition au bruit et seuils à 4 kHz

Les sujets n'ont aucune histoire d'antécédents personnels ou très peu. Par contre les loisirs ont engendré un Eₐ moyen de 89 dBa (59,3) pour les élèves du secteur académique et de 91,6 dBa (54,5) pour ceux du professionnel. La plus grande proportion d'élèves du professionnel pour chaque catégorie de d'âges à l'âge. Les élèves de l'académique rapportent proportionnellement moins d'expérience de travail. Elles semblent plus bruyantes mais de moins longue durée. Le Eₐ lié au travail serait donc plus élevé chez les élèves du professionnel.

Le Eₐ lié à l'apprentissage scolaire varie de 74 à 98 dBa (près de 35% des sujets ont des Nₐ de 80 dBa et moins de 20%). Il y a une grande variation entre les différents programmes, les niveaux d'un même programme et les écoles.

L'analyse de variance a été utilisée pour l'étude de variations des seuils à 4 kHz. Les résultats sont présentés dans le tableau 2. Les antécédents de loisirs et les catégories de nombre de carabouches utilisées. L'exposition en cours d'apprentissage est définie de différentes façons, soit le secteur d'apprentissage (professionnel, académique), le profil d'études, le programme d'apprentissage et la catégorie d'exposition au bruit cumulée en cours d'apprentissage (Eₐ). À 4 kHz, on observe une différence statistiquement significative (p = 0,02) de 1 dBHL entre les seuils auditifs des élèves du professionnel et ceux des élèves de l'académique (tableau 2), les premiers ayant le seuil moyen le plus élevé. Dans un intervalle de confiance de 95%, la différence se situe entre 0,2 et 1,8 dBHL.

DISCUSSION

Bien que la différence soit faible (1 dBHL) entre l'audition des élèves des deux secteurs d'apprentissage, elle apparaît être systématique pour les différentes percentiles de la distribution des seuils. Nous ne pouvons pas isoler un indice de risque plus spécifique à l'exposition: par exemple, le programme d'études, la dose d'énergie cumulée (Eₐ scolaire) ou le mode temporel d'exposition.

La dose d'énergie cumulée (Eₐ), indice dérivé du principe de l'égalité d'énergie, bien que statistiquement significative dans l'explication de la variation des seuils, ne montre pas de tendance nette entre le Eₐ scolaire et les seuils à 4 kHz. Il est peu probable que les mesures comme celles soient au cause. Les individus examinés pourraient se situer dans une zone de progression lente de l'effet du bruit sur l'audition. Cela expliquerait qu'il n'y ait pas de relation claire entre la dose d'exposition spécifique (Eₐ scolaire) et les seuils auditifs.

Il se pourrait ainsi que l'indice de risque soit davantage lié au mode temporel d'exposition qu'à la dose cumulée. Pour une même valeur de Eₐ et peut-être même pour des valeurs différentes, le mode temporel d'exposition pourrait déterminer le risque d'atteinte. L'influence de la variable profil d'apprentissage est en dehors du seuil de la signification statistique (p = 0,06) et les résultats ont tendance à associer une augmentation des seuils à une augmentation du temps total d'exposition. Il y aurait lieu d'in-
vastiquer plus à fond le mode d'exposition des élèves en spécifiant, par exemple, les périodes de production intensive et la nature des profils dits marginaux.

L'hypothèse selon laquelle le mode temporel d'exposition influencerait le risque d'atteinte réfère au mode de récupération d'un décalage temporel du seuil auditif comme indice d'acquisition d'un décalage permanent. (4) Une période de production intensive (fin d'un projet, période d'examen, stage) surtout si elle s'accompagne d'une exposition au bruit extérieur à l'apprentissage (loisirs, travail) pourrait représenter une exposition excessive pour les élèves du professionnel et, en agissant sur le mode de récupération, augmenter le risque d'atteinte permanente.

Une étude de type cas-témoins contribuerait à cerner de façon plus explicite le facteur de risque associé au fait d'être étudiant dans un secteur professionnel.

Le nombre de sujets examinés et la procédure d'examen permettent, croyons-nous, de se baser sur ces résultats comme première valeur de référence spécifique dans la description des seuils auditifs de jeunes québécois de ces âges. En adoptant le principe définissant une perte auditive significative comme étant la valeur du seuil au 90e percentile à une fréquence donnée, nous obtenons pour les sujets du secteur académique des valeurs 10 dB HL à 0,5 kHz, de 8 dB HL à 1, 2, 3 et 4 kHz et de 15 dB HL à 6 kHz. Ces valeurs diffèrent légèrement de celles proposées par ISO 1979-1986 (différence de -3 à +3 dB HL). Il y aurait lieu de compléter l'étude des seuils d'audition de la population canadienne pour les autres groupes d'âges de chaque sexe.

D'un point de vue pratique, les résultats impliquent qu'un effort de prévention du risque d'atteinte auditive soit concrétisé au niveau de la formation scolaire. À cause du caractère particulier de leur situation d'apprentissage, de nouveaux moyens sont à développer pour réduire le risque. Ainsi l'emploi de protecteurs auditifs n'est sans doute pas indiqué pour des élèves et des enseignants qui, dans cette situation d'éducation, doivent forcément communiquer verbalement.

D'autre part, il y a raison de croire que les élèves présentent des signes de fatigue auditive (5) et qu'ainsi leur aptitude à suivre un cours théorique après une période en atelier peut être sérieusement affectée. (6) Il serait intéressant d'étudier ce phénomène.

La prévention supposée au niveau des ateliers professionnels devrait, à plus forte raison, s'appliquer aussi aux enseignants exposés de façon plus marquée.

De plus, étant donné l'exposition au bruit généré par les loisirs et le travail, la sensibilisation au risque d'atteinte auditive devrait s'étendre aussi bien aux élèves du secteur académique qu'aux élèves professionnels.

BIBLIOGRAPHIE


Tableau 1: Sensibilité auditique (moyennes des deux oreilles) par fréquence. Comparaison de deux études.

<table>
<thead>
<tr>
<th>Fréquence (Hz)</th>
<th>Sensibilité de référence (dB)</th>
<th>Sensibilité de référence (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>4,70</td>
<td>4,70</td>
</tr>
<tr>
<td>1000</td>
<td>1,82</td>
<td>1,45</td>
</tr>
<tr>
<td>2000</td>
<td>1,08</td>
<td>1,14</td>
</tr>
<tr>
<td>3000</td>
<td>2,54</td>
<td>2,75</td>
</tr>
<tr>
<td>4000</td>
<td>2,10</td>
<td>2,10</td>
</tr>
<tr>
<td>6000</td>
<td>7,50</td>
<td>7,65</td>
</tr>
</tbody>
</table>

Tableau 2: Sensibilité d'audition au niveau du secteur d'orientation et les résultats d'études professionnelles.

<table>
<thead>
<tr>
<th>Domicile et profil</th>
<th>n</th>
<th>X</th>
<th>s</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Académique</td>
<td>370</td>
<td>1,91</td>
<td>3,8</td>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td>Professionnel</td>
<td>277</td>
<td>2,91</td>
<td>5,4</td>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td>Prise en considération des deux types d'enseignement</td>
<td>174</td>
<td>2,50</td>
<td>5,5</td>
<td>10</td>
<td>55</td>
</tr>
<tr>
<td>Prise en considération des deux types d'enseignement</td>
<td>43</td>
<td>2,53</td>
<td>3,5</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Prise en considération des deux types d'enseignement</td>
<td>88</td>
<td>3,85</td>
<td>6,1</td>
<td>10</td>
<td>18</td>
</tr>
<tr>
<td>Prise en considération des deux types d'enseignement</td>
<td>285</td>
<td>2,09</td>
<td>3,7</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Prise en considération des deux types d'enseignement</td>
<td>27</td>
<td>2,70</td>
<td>6,8</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Prise en considération des deux types d'enseignement</td>
<td>15</td>
<td>3,80</td>
<td>6,0</td>
<td>5</td>
<td>15</td>
</tr>
</tbody>
</table>

1. Descriptive analysis of a new form of base: 2 mos, 2 mos
2. The formation of base and of a model of site, 2 mos
3. Descriptive analysis of a new form of base: 2 mos, 2 mos
4. Prise en considération des deux types d'enseignement: 2 mos, 2 mos
5. Prise en considération des deux types d'enseignement: 2 mos, 2 mos
6. Prise en considération des deux types d'enseignement: 2 mos, 2 mos
7. Prise en considération des deux types d'enseignement: 2 mos, 2 mos
8. Prise en considération des deux types d'enseignement: 2 mos, 2 mos
9. Prise en considération des deux types d'enseignement: 2 mos, 2 mos
DESCRIPTIVE ANALYSIS OF HEARING LOSSES IN CHILDREN EXPOSED BEFORE BIRTH TO INDUSTRIAL NOISE

N.M. Lalande, R. Hétu and J. Lambert

aÉcole d'orthophonie et d'audiologie et bDépartement de médecine sociale et préventive, Université de Montréal, C.P. 6128, Succursale A, Montréal, Canada H3C 3J7

Several investigations have shown a relatively high prevalence of sensorineural high frequency hearing loss in children. The reasons usually put forward to explain such losses were either the experience of noisy activities (Weber et al., 1967; Cosad et al., 1974; Byrnsen and Byrnsen, 1981; Shepard et al., 1983) or a genetic defect (Harr et al., 1983; Fisch, 1981); they relied on the facts that boys were more often affected than girls and that the prevalence of high tone losses among children increases with age. It was assumed a) that boys are more frequently submitted to loud recreational noises when compared to girls and that the chances of being exposed to loud noises increase with age or b) that the genetic defect would be sex-linked. However, the causal factors called upon by these investigators do not explain the high frequency loss found in young girls.

A recent exploratory study has shown a statistically significant association between the cumulated acoustic energy received during pre-natal life and the hearing threshold for the worst ear at 4000 Hz (Lalande and Hétu, 1985). The aim of this paper is to describe the hearing losses found for those children and the practical implications of having a mild permanent hearing loss early in life.

METHODOLGY

A total of 167 children had their hearing tested by conventional air-conduction audiometry (Maico MA-41 or Grason Stadler 1703B) and by tympanometry (Madsen 25 330). Of these, 31 children were excluded from the analyses. The final sample included 63 girls and 68 boys approximately uniformly distributed from 4 to 7 years of age, with slightly younger boys from 8 to 10 years of age. Reasons for exclusion were: invalid hearing test (n=3), bilateral middle ear problem at the time of the test (n=21) and positive answers during the interview to a known cause of hearing loss (n=12).

The mothers of these children originated from different sectors of industrial activity (n=14) and plants (n=45), implying that the selected children have been exposed to various noise conditions before birth. The details of the selection criteria for plants, mothers and children have been described in Lalande et al., (1985).

Besides the measurement of hearing thresholds from 500 to 8000 Hz, the occupational history of the mother during pregnancy and the auditory history of the child and the mother were collected via a questionnaire-interview. The noise measurements at job sites occupied during pregnancy were obtained with an integrating sound level meter (Bruel & Kjaer 2218 with an ½ inch microphone). From the noise levels and their corresponding duration of exposure for a typical working day, a representative daily noise dose was computed for each job position held during pregnancy (LPEQ). Then, week (LPEQ-7) and month pregnancy doses were calculated (LPEQ-30). The independent variable was the LPEQ-9m. Three classes of noise exposure with relatively similar sample sizes were considered, namely LPEQ-9m of 65 to 75, 75 to 85 and 85 to 95 dB, thus defining 3 groups of children. Concomitant variables (n=18) that could have been related to the pre-natal noise exposure and/or to the hearing thresholds were taken into account in the analysis of the data (Lalande et al., 1985).

RESULTS

As shown in Table 1, when the worst hearing threshold for each frequency tested is considered, the mean values for 3 classes of noise exposure are relatively similar for frequencies equal to and lower than 2000 Hz. Above this frequency, the means tend to increase with the frequency and the exposure. An effect of noise exposure was obtained with an analysis of variance at 3000 Hz (p<0.05). Moreover, a Scheffe test revealed a significant difference between the mean for the higher class of noise exposure and the means for the lower classes (p<0.05). At 6000 Hz, a similar tendency was found (anova; p<0.08). The same pattern of results was obtained for the left and for the right ear. These findings are at variance with the well known effect of excessive noise exposure after birth (Melnick, 1978).

Table 1. Mean values (dB HL re: ANSI 1969) and standard deviations (in parentheses) of the worst hearing threshold for each frequency tested and for 3 classes of noise exposure (LPEQ-9m) before birth (N=131).

<table>
<thead>
<tr>
<th>LPEQ-9m (dB)</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>3000</th>
<th>4000</th>
<th>6000</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td>65-75</td>
<td>3.9</td>
<td>2.2</td>
<td>1.2</td>
<td>3.0</td>
<td>3.7</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>(3.7)</td>
<td>(4.8)</td>
<td>(5.9)</td>
<td>(5.2)</td>
<td>(6.2)</td>
<td>(7.9)</td>
<td>(9.5)</td>
<td></td>
</tr>
<tr>
<td>75-85</td>
<td>4.7</td>
<td>2.8</td>
<td>0.2</td>
<td>0.6</td>
<td>2.9</td>
<td>6.1</td>
<td>10.9</td>
</tr>
<tr>
<td>(3.6)</td>
<td>(4.6)</td>
<td>(5.5)</td>
<td>(5.3)</td>
<td>(6.3)</td>
<td>(7.7)</td>
<td>(8.8)</td>
<td></td>
</tr>
<tr>
<td>85-95</td>
<td>3.3</td>
<td>2.1</td>
<td>1.4</td>
<td>3.4</td>
<td>5.6</td>
<td>7.9</td>
<td>12.3</td>
</tr>
<tr>
<td>(3.2)</td>
<td>(4.4)</td>
<td>(4.9)</td>
<td>(5.8)</td>
<td>(7.1)</td>
<td>(6.6)</td>
<td>(9.9)</td>
<td></td>
</tr>
</tbody>
</table>

Since the mean thresholds are better than 13 dB HL at all frequencies tested, it could be thought that the hearing losses found in 31% of the children were rather mild and had no adverse effect on them. However, to better understand the extent of the hearing losses, one should also look at the number of frequencies affected for each child. Since for 29 out of 131 children, their hearing thresholds were considered for only one ear, the distribution of pure tone hearing losses per ear and number of frequencies are presented for 102 children. Criteria for a significant hearing loss (above the 90th percentile for children with normal hearing) was a threshold greater than 10 dB HL from 1000 to 4000 Hz, 15 dB HL at 500 and 6000 Hz and 20 dB HL at 8000 Hz.

The last column of Table II indicates that 31.4% of the children had a significant hearing loss. For 9.8%, it involved only one ear and one frequency, but for the other 21.6%, there was a hearing loss at more than one frequency. Thus, even if the hearing deficit was mild, 69% of the affected children had a loss at two or more audiometric frequencies. Now, it is known that in presence of acoustic trauma, there is not only a reduction of auditory sensitivity but also other auditory impairments. Moreover, for some of these impairments (ex: decreased frequency selectivity), it affects more frequencies than those for which a loss of sensitivity is observed (Salvi et al., 1982).
TABLE II. Proportion (%) of children with normal hearing and with a hearing loss for 3 classes of pre-natal exposure to noise.

<table>
<thead>
<tr>
<th>Hearing Status</th>
<th>$L_{eq-9m}$ (dB)</th>
<th>All Doses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>65-75</td>
<td>75-85</td>
</tr>
<tr>
<td>Normal hearing</td>
<td>79.2</td>
<td>67.7</td>
</tr>
<tr>
<td>Hearing loss/1 ear</td>
<td>12.5</td>
<td>22.6</td>
</tr>
<tr>
<td>1 frequency</td>
<td>12.5</td>
<td>9.7</td>
</tr>
<tr>
<td>2 frequencies</td>
<td>-</td>
<td>12.9</td>
</tr>
<tr>
<td>3 frequencies</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hearing loss/2 ears</td>
<td>8.3</td>
<td>9.7</td>
</tr>
<tr>
<td>1 frequency</td>
<td>6.2</td>
<td>-</td>
</tr>
<tr>
<td>2 frequencies</td>
<td>-</td>
<td>6.5</td>
</tr>
<tr>
<td>3 frequencies</td>
<td>4.1</td>
<td>3.2</td>
</tr>
<tr>
<td>4 frequencies</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total (%)</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Total (no/children)</td>
<td>n=24</td>
<td>n=31</td>
</tr>
</tbody>
</table>

Concerning a possible dose-effect relationship, it is interesting to note in Table II the distribution of children with normal hearing and with a hearing loss for the 3 classes of noise exposure before birth. There is a tendency for the proportion of children with normal hearing to decrease when they have been exposed to higher noise doses. Conversely, the proportion of children with a bilateral hearing loss increases with the noise dose.

DISCUSSION

Theoretical Implication

The hearing losses found in children which could be attributed to their prenatal exposure to noise were at variance with those generally met in teenagers and adults submitted to excessive noise doses. In most cases, above 2000 Hz, the hearing loss increases with frequency and the usual 4000 or 6000 Hz dip was not present. So, the present results suggest that the mechanism of impairment is probably different when the noise is transmitted through the abdominal wall and the amniotic fluid of the mother rather than through air. For a better understanding of the mechanism responsible for the prenatal hearing damage, as well as the site of the lesion in the auditory system, laboratory studies with an animal model should be carried out.

Practical Implications

Except in one case out of 31, the mothers did not suspect the presence of a hearing loss in their child. Consequently, a) screening for mild high frequency hearing losses in children should not rely on questionnaires and interview to parents and b) primary prevention should be implemented to avoid any permanent hearing damage to the foetus. The present results suggest that, temporarily, the exposure limit to noise during pregnancy should not exceed 85 dBA-8h, assuming that most pregnant workers are at their job during 7 to 8 months of their pregnancy.

The need of preventing any damage to the auditory system of the foetus is supported by the fact that the consequences of a mild hearing loss are very likely different for young children when compared to adults, particularly when it is not suspected. This presumption is supported by data from the present study. Namely, the risk of having a learning problem at school (as reported by the mother) was a) fourfold higher (3 out of 14 or 21.4% versus 4 out of 82 or 4.8%) when the children had a hearing loss at 4000 Hz as compared to normal hearing ($X^2=4.9; p=0.03$) and b) two fold higher (4 out of 11 or 36.4% versus 3 out of 20 or 15%) when the children had a bilateral hearing loss rather than a unilateral hearing loss ($X^2=20.9; p<0.001$).

ACKNOWLEDGMENTS

This study was financially supported by Health and Welfare Canada. The statistical analyses were done by Claudine Legault and the English manuscript was read by John Dudley.

REFERENCES


AETIOLOGICAL MODELS OF HEARING LOSS IN THE WORKPLACE

R. Phaneuf, R. Hétu
GAUM, Université de Montréal, Pavillon Marguerite d’Youville, C.P. 6128, Succursale A, Montréal, Québec Canada, H3C 3J7

Noise induced hearing loss is the most prevalent irreversible industrial disease (1). Estimates of noise induced hearing loss range between 8 to 12 per thousand persons (2,3). Though the relationship between noise and the audiometric thresholds has received considerable attention, (4) we suggest that there may be other factors in the occupational environment that may account for hearing loss. Indeed Aitherley and Noble (1985) have reported investigations dating to the nineteenth century, which speculated that other factors in the workplace might contribute to hearing loss (5).

In figure 1, we propose a model for hearing loss with two major aetiological components. The first is associated with the workplace. The second is composed of the more familiar expressions of presbycusis, sociacusis, and neocacusis which are not associated with the workplace.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Occu. occupational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occupational Noise</td>
<td>Noise</td>
</tr>
<tr>
<td>Respiratory irritants</td>
<td>Disease</td>
</tr>
<tr>
<td>Oto-toxic exposures</td>
<td>Oto-toxic drugs</td>
</tr>
<tr>
<td>Trauma</td>
<td>Trauma</td>
</tr>
</tbody>
</table>

Figure 1. An Aetiological Model for Hearing loss in an industrial population.

We shall discuss three possible models from this paradigm of expected hearing loss at 4000 Hz. In each model we assume that there is a continuous working exposure of 90 dBA for age 20 until age 65.

1. THE JOINT EFFECT OF GUNFIRE AND NOISE

The combined effect of shooting in addition to noise is examined by reviewing data from a NIOSH study of coal miners whose exposures were estimated equivalent to 90 to 95 dBA-8h and who had shot 50 rounds/year for six years or more, or 1000 rounds/year or more for three years (6). These were compared with workers who had never been exposed to gun fire. The pattern of hearing loss in the two groups is similar up until about 35 years of age. Between 55 and 64 years of age, the heavy shooters have comparatively a decreased hearing sensitivity. This difference diminishes from age 45 until it no longer exists by age 65. This suggests that the additional effect of gunfire exposure makes little if any difference in the final outcome. In figure 2, we propose an empirically derived model to demonstrate the observed plateauing of the combined effect of gunfire in addition to noise. The curve is portrayed against the expected results of a continuous exposure to 90 dBA-8h. This model conciliates data that showed an increased average effect for hearing loss attributable to gunfire (7), and an observed non-difference seen in two groups of foundry workers of ages 60 to 64 (8). Based on the above data it can be inferred that gunfire has little meaning when considering the average compensation cases since, by the time the average hearing threshold reaches the low 25th percentile of 25 dB at 90 dBA-8h, the average hearing level for 500, 1000, and 2000 Hz, there would likely be no systematic difference associated with gunfire.

2. INDUSTRIAL DISEASES AND HEARING LOSS

The prevalence of middle ear disorders found among industrial workers often surpasses that which is found in randomly selected samples (6,9). The excess prevalence may be found in work environments involving exposures to irritants: dusts and volatile chemicals that can cause an allergic swelling of the Eustachian tube leading to middle ear disorders (MED) (10). In a Quebec foundry, we found 12 (75%) workers exposed to acid fumes, whereas the proportions of 212 workers without exposure to irritants was 19%. Though we will not contend that this difference in the proportions is a conclusive proof as the cause of occupational middle ear disorders, we feel sufficiently confident to put forward its plausibility.

We propose an empirical model of an additive effect of hearing loss of 13 dB at all ages at 4000 Hz attributable to middle ear disorders of occupational origin. This is founded on a mean difference observed of 13 dB found in young men in the original "Hearing and Noise in Industry" study (4), and for seasoned foundry workers of ages 60 to 64 (7).

3. OCCUPATIONAL AND OTO-TOXIC EXPOSURES

Given that there are occupational exposures that on to themselves can cause a loss of hearing sensitivity such as lead and carbon monoxide given enough exposure (11, 12), we can advance three possible aetiological models for three combined effects with

Figure 2. A model simulating the effect of gunfire in addition to industrial noise.

Figure 3. The proposed effect of an occupational middle ear disorder and noise exposure at 4000 Hz in comparison to noise exposure only.
DIFFUSION

The implications of the above aetiological models suggest that working conditions other than noise can contribute to a person's hearing loss. The consequence of this inference relates to the noise exposure limits. The current noise protection criteria does not assume that individuals may have simultaneous work-related ototoxic exposures or might have an excess risk to middle ear dysfunctions associated with work exposures.

One possible course of action is to recommend more stringent noise level requirements in those jobs where there is a greater risk of exposure to upper respiratory irritants and ototoxic. These same issues bear on an individual's compensation claim.

REFERENCES


ACOUSTIC IMPEDANCE MEASUREMENTS OF CHILDREN'S EAR

K.Okabe, S.Kanmu, H.Hamada, T.Miura* and H.Funai**

* Tokyo Denki University
2-2 Kanda nishiki-cho, Chiyoda-ku, Tokyo 101 Japan.
** Department of Otolaryngology, Branch Hospital, University of Tokyo
3-28-6 Mejirodai, Bunkyo-ku, Tokyo 112 Japan.

1. INTRODUCTION

Acoustic impedance measurements of human ears provide important information on diagnosis of middle ears and fundamental data for the design of earphone calibration couplers. Many authors (for example, see [1]-[4]) measured the acoustic impedance on adult subjects.

Unfortunately, however, there exists no published data on complex ear canal input impedance of children. In order to fulfill the lack of the data and develop an ear simulator, we made measurements with about 100 children. The subjects were aged from 3 years to 12 years and all had normal hearing.

We utilized an acoustic tube and measured acoustical transfer function in the tube terminated by the human ear canal. The sound source was broad band random noise. Cross-power spectrum method [5] for FIR (finite impulse response) system identification was applied to estimate the transfer function.

The results of the impedance measurements and estimates at the eardrum are discussed.

2. METHOD

2.1 Acoustical impedance measurements

The principle of the measurements is depicted in Fig.1. We assume plane wave propagation in the measuring tube with a constant cross-section and losses are negligible. In the figure, incidental wave propagates from left to right.

The complex reflection coefficient \( R(\omega) \) at the reference plane X can be expressed as eq.(1).

\[
R(\omega) = \frac{1-H(\omega)\exp(j\omega t_1)}{1-H(\omega)\exp(-j\omega t_2)}
\]

\[
t_1 = 2(l_1 - l_2)/c, \quad t_2 = l_2/c
\]

where \( H(\omega) \) denotes the acoustical transfer function from the plane A to B, and c is velocity of sound.

The acoustical impedance \( Z(\omega) \) at the reference plane X can be calculated from the well-known relationship of eq.(2).

\[
R(\omega) = \frac{Z(\omega) - Z_0}{Z(\omega) + Z_0}
\]

In our measurements, we chose 4.76 cm for \( l_1 \) and 1.40 cm for \( l_2 \).

2.2 Estimation of ear canal dimensions

Method for reconstructing boundary conditions from reflected waves is known as the inverse Sturm-Liouville problem. In many fields such as quantum mechanics, electromagnetics, geophysics and speech research, the inverse methods were successfully applied [6].

In our context, the ear canal must be reconstructed from reflections, which is fitted with the terminal end of the measuring tube (Fig.2). Hudde [4] is the first author to take into account the effects of the ear canal shape on the estimation of the eardrum impedance. He used three probe tube microphones to measure acoustical transfer functions in the ear canal and calculated its area function and the impedance at the eardrum.

We utilized the Goupilleau solution cited in Ware and Aki [7] to solve the
inverse problem, i.e. the reconstruction of the ear canal cross-sections from acoustical data. See the reference for derivations of the detailed algorithm. This method requires reflections as a function of time against impulsive incidence. These reflections can be obtained from \( R(w) \) by calculating its inverse Fourier transform. Reconstructed ear canal dimensions were used to estimate the impedance at the eardrum.

3. RESULTS

Before the measurement, our subjects were clinically assessed to have no disease in the middle ear. Acoustical data of 1 sec was sufficiently long to estimate the transfer function in the measuring tube. As an example of the results, Fig. 3 shows the average impedance measured at the mid point in the ear canals of 5 year-old children (9 males and 4 females).

From the measured transfer function \( H(w) \) on each subject, we calculated the ear canal dimensions using eq.(1) and the algorithm in ref.[7]. The eardrum impedance was estimated by using the ear canal input impedance and the ear canal dimensions. The averaged eardrum impedance of the 13 subjects is shown in Fig.4.

4. DISCUSSIONS

The comparison of our results with Morton and Jones [8] in Fig.3, especially below 1 kHz, shows that children have relatively high input reactance than adults. This suggests the small ear canal and small middle ear cavity of children.

The mean radius of 0.281 cm was calculated from the reconstructions of the residual ear canal cross-sections. The value is apparently smaller than 0.375 cm of adults [1]. Estimates of eardrum impedance, especially reactance below 1 kHz in Fig.4, have larger values than adults.

These findings provide important information to construct the ear simulator of children. Detailed studies on the impedance difference between male and female or on the difference among the age-classes were carried out for the development of the ear simulator.

REFERENCES


Fig.3 Mean value of ear canal input impedance measured on 13 children of 5 years old (9 males and 4 females). The dotted lines indicate the range of one standard deviation from the mean values. The broken lines denote the results on 19 male adults measured at the tip of an insert earphone (Morton and Jones, 1956).

Fig.4 Mean value of eardrum impedance of 13 children of 5 years old (9 males and 4 females). The dotted lines indicate the range of one standard deviation from the mean values. The broken lines show the eardrum impedance of 10 male adults (Zwischen, 1970).
DYNAMIQUE DU REFLEXE ACOUTIQUE EN FONCTION DES PARAMETRES DE LA STIMULATION ACOUTIQUE CHEZ L'HOMME

D. Luthi, P. Avan, C. Mempy, F. Bisaro, M. Teysou

Service Central de Biophysique et de Médecine Nucléaire (Pr. M. HURGET) Hospital Lariboisière, 2 rue Ambrine Paris, 75010 Paris, France

L'étude du réflexe stapedien chez l'homme s'est largement développée ces dernières années, tant sur le plan de la recherche fondamentale (1,8) que sur celui du diagnostic clinique (6). Malheureusement, on constate dans les études fines des caractéristiques du réflexe une grande dispersion des résultats dûe à 2 causes principales :

- la variabilité interindividuelle importante des réponses,
- l'hétérogénéité des protocoles expérimentaux et du choix des définitions des paramètres à mesurer.

Le temps de latence du réflexe est un exemple typique : sur le plan physiologique, c'est le délai qui s'écoule entre l'arrivée du stimulus acoustique efficace au récepteur périphérique et la réponse des muscles de l'oreille moyenne. En pratique, 2 types de critères sont retenus pour préciser le début de la réponse réflexe sur un enregistrement :

- certains auteurs (2,7) déterminent directement l'instant où une première déflexion est détectable, et se heurtent alors notamment à des problèmes de sensibilité de la détection ;
- d'autres (1,3) repèrent l'instant où la déflexion atteint un pourcentage donné, faible (5 ou 10 %) de sa valeur maximale. La valeur ainsi déterminée est plus stable mais systématiquement supérieure au véritable temps de latence, puisqu'en ajoutant à ce dernier une fraction du temps total d'établissement du réflec-
- xe, qui dépend aussi du temps de réponse de l'appareil ouvrage, notamment en impédancemétrie.

Dans ce travail, nous avons choisi de définir le temps de latence selon le 2ème critère, c'est à dire comme étant la durée séparant le déclenchement du stimulus de l'instant où la réponse atteint 10 % de son amplitude maximale. Le but est l'étude des variations du réflexe stapedien selon les divers stimulus acoustiques utilisés, sous l'angle de la protection éventuelle apportée par le réflexe ; il n'est pas gênant dans ce cas d'obtenir un résultat par excès. Par con-
- tre, lorsqu'on étudie l'influence d'un paramètre du stimulus acoustique sur une caractéristique physiologique du réflexe, il est important de savoir si la dépendance observée est réelle ou résulte des biais introduits par les définitions choisies.

PROTOCLE EXPERIMENTAL :

Nous avons donc effectué sur 20 sujets normaux une étude systématique, par impédancemétrie, du rôle des paramètres de la stimulation sur le temps de latence du réflexe, et la durée d'établissement du réflexe.

Les stimulations acoustiques utilisées ici ont les caractéristiques suivantes :

- enveloppe trapezoïdale de temps de montée (Tm) et de descente égaux et de valeur 0,1, 50 et 300 ms et de durée de plateau (Tp) égale à 1 ou 10 s,
- contenu spectral pouvant être soit un ton pur de fréquence (f) 500 ou 4000 Hz, soit un bruit blanc filtré entre 50 et 10000 Hz (pentes de courbe 48 dB/octave).

La chaîne de stimulation acoustique possède une dynamique de 115 dB et une courbe de réponse plate à 1 db près dans la gamme de fréquence utilisée ; la distorsion est négligeable jusqu'à 112 dB SPL.

Un ordinateur Digital Minicomputateur numérique des signaux de variation d'impédance en provenance d'un impédancemètre Madison 20 75 a-

avec une résolution temporelle de 4,5 ms pour un stimulus de durée de plateau de 1 ms. Un filtrage digital peut être utilisé pour réduire le bruit d'enregistrement.

Protocole de traitement du signal :

Le maximum d'amplitude du réflexe est automatiquement déterminé ainsi que les principales durées caractéristiques de son déroulement :

- le temps de latence du réflexe, défini plus haut : L10
- les autres instants remarquables décrivant la dynamique du réflexe qui, par extension, sont définis de manière analogique à la latence en se référant aux instants où l'amplitude de la réponse atteint 10 % de l'amplitude maximale à la montée (L66+), et 66 et 10 % de l'amplitude maximale lors de la phase de décru-

sance à l'arrêt de la stimulation (L66-, L10-).
- les logiciels de traitement usuellement utilisés permettent de s'affranchir des principaux artefacts d'enregistrement, tels que derive de la ligne de base, effets de rebond transitoire, etc.

Chaque stimulus est appliqué dans l'oreille contrala-

térale par rapport aux mesures impédancemétriques ; on détermine d'abord le seuil auditif pour le stimulus considéré puis le seuil du réflexe acoustique (recherche ascendante par pas de 1 dB, le niveau du seuil étant défini comme le niveau nécessaire pour produire une déviation reproductible le plus proche de 0,5 V à 1 dB près en sortie de l'impédancemètre). Chaque type de stimulus est ensuite présenté au sujet avec des intensités sonores croissantes par pas de 2 ou 5 dB jusqu'à 15 dB au moins au dessus du seuil.

RESULTATS ET DISCUSSION :

On effectue tout d'abord, chez chaque individu testé, une détermination précise du niveau du seuil du réflexe, et toutes les intensités sonores sont en-

suite exprimées par référence à ce niveau, (donc no-

tées en dB re S.P.L). Ce niveau dépend essentiellement du contenu spectral du stimulus : il est plus élevé d'environ 20 dB pour un ton pur (moyenne 66 dB HL (a=9) pour f=500 Hz, 92 dB HL (a=9,5) à f=4000 Hz), qu'un bruit blanc (moyenne 64 dB HL (a=10) (5). L'amplitude maximale du réflexe acoustique est un paramètre difficile à standardiser car la morphologie des courbes amplitude/temps présente de grandes variations individuelles (croissance linéaire de l'amplitude jusqu'au niveau sonore maximal admisible (environ 30 dB au dessus du seuil) chez la plupart des sujets, inflation de la courbe ou apparition d'un plateau au delà de 15 dB au dessus du seuil chez les autres), (le contenu spectral du stimulus influe également sur les amplitudes attein-

- tes pour une intensité sonore donnée (tableau 1) ; par contre, dans les limites de nos paramètres, l'amplitude du réflexe ne dépend pas significativement du temps de montée du stimulus (tabl. 1).

<table>
<thead>
<tr>
<th>TM(ms)</th>
<th>50</th>
<th>100</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplitude maximale</td>
<td>3,45(1,8)</td>
<td>3,77(1,4)</td>
<td>5,02(1,7)</td>
</tr>
<tr>
<td>(f=500Hz)</td>
<td>(v)</td>
<td>2,57(1,3)</td>
<td>2,60(1,3)</td>
</tr>
<tr>
<td>6000Hz</td>
<td>2,44(1,5)</td>
<td>2,50(1,3)</td>
<td>2,54(1,3)</td>
</tr>
</tbody>
</table>
Pour chaque sujet et chaque type de stimulus, le temps de latence dépend tout d’abord de l’intensité sonore : c’est une fonction rapidement décroissante de l’intensité qui tend vers une limite inférieure (latence minimale) atteinte pour des niveaux de l’ordre de 10 à 15 dB re S.Px., même chez les sujets pour lesquels la courbe amplitude/intensité continue à croître linéairement à ce niveau (Z/7). La valeur de cette limite inférieure est portée sur la figure 1 ; sur cette figure, l’écart-type mesuré est d’origine statistique ; il est important, illustrant la variabilité inter individuelle de la latence. La variabilité intra individuelle est par contre beaucoup plus limitée.

Une double dépendance apparaît :
- pour un temps de latence mesuré dépend du contenu spectral du stimulus : il est minimal pour un son pur de 500 Hz, nettement plus élevé pour un son pur de 4000 Hz ou un bruit blanc ;
- pour chaque fréquence utilisée, la latence minimale apparaît liée de manière linéaire au temps de montée, au moins dans la gamme étudiée (0-300 ms), avec une pente nettement inférieure à 1 (de l’ordre de 0,33).

La première dépendance semble paradoxale sur le plan physiologique. En fait, son origine est bien mise en évidence lorsqu’on étudie parallèlement la durée de la phase d’établissement du réflexe décrite par le paramètre L(66%)=L(10%) représenté sur la figure 2 : la durée de cet intervalle est fortement dépendante de la fréquence du stimulus (d’autant plus que L(66%)<L(10%) en représente une valeur biaisée par défaut). La "latence" L(10%) contient un pourcentage de cette durée, et la dépendance qu’elle manifeste en fonction de la fréquence résulte donc essentiellement de cette dépendance. Il est probable également que l’absence de diminution de L(10%) au delà de 15 dB re S.Px., alors que l’amplitude du réflexe continue généralement à croître, provient d’un phénomène analogue (l’amplitude augmentant, le temps nécessaire pour atteindre 30 % de celle-ci augmentant, la latence mesurée est artificiellement allongée).

Par contre, on constate que ni l’amplitude (Tab. 1), ni la durée du temps de montée de la réponse tel que nous l’avons défini ne dépendent significativement du temps de montée du stimulus : l’effet de ce dernier sur la latence n’est donc pas sujet à caution.

CONCLUSION

2 faits intéressants sont donc mis en évidence dans la limite des valeurs étudiées au moins :
- la latence dépend linéairement du temps de montée du stimulus (7), avec une pente très inférieure à 1, donc une augmentation importante de ce dernier ne se répercute pas complètement sur la latence, vraisemblablement en relation avec le phénomène de sommation temporelle (4) qui contribue à accélérer le déclenchement du réflexe ;
- la durée d’établissement du réflexe est apparemment relativement indépendante du temps de montée du stimulus.

BIBLIOGRAPHIE


![fig. 1 : Latence minimale en fonction du temps de montée du stimulus pour les 3 contenus spectraux étudiés (moyenne sur 20 sujets).](image1)

![fig. 2 : Durée d’établissement du réflexe en fonction du temps de montée du stimulus pour les 3 fréquences étudiées (moyenne sur 20 sujets).](image2)
A THREE-DIMENSIONAL MEASUREMENT OF HUMAN AURICLE

FOR THE PURPOSE OF DUMMYHEAD CONSTRUCTION

Kinotoshi Fukudome and Michihiro Matsuoto

Dept. of Acoustic Design, Kyushu Institute of Design, 4-9-1 Shiobaru Minami-ku, Fukuoka, 815 Japan

INTRODUCTION

Let us consider the realization of the dummyhead-headphone system for many listeners; Through this system, each of them is able to hear the same sound as he would hear in the sound field where the dummyhead is located.

An exact reproduction of the sound pressure on the eardrum of the listener in the standard original sound field can be made if the shape of the dummyhead is identical to that of the listener's head except the inner part from its eardrum entrance. There are differences between individuals about the shape and size of the head and auricle. For many listeners, therefore, the exact reproduction of the eardrum pressure would not be expected. However, the ear is not able to discriminate small differences of stimuli within threshold. This enables a listener to hear the same sound as in the original sound field if the differences of the sound generated through this dummyhead-headphone system is within threshold. If such a dummyhead commonly available for each listener, is defined as the standard dummyhead.

This paper is one of studies aiming to determine the standard dummyhead. Already the study on a three-dimensional measurement of human head was made to determine the specification of the head shape without the auricle, the nose and hair. In this paper, a method which is more suitable for dummyhead construction is presented of describing numerically the three-dimensional shape of human auricle.

MEASURES DESCRIBING THE SHAPE OF HUMAN AURICLE

Human auricle is complicated in shape as shown in Fig.1. (The crus-intersection, the fiducial concha line, etc with * in Fig.1 are defined by the authors. See Appendix.)

![Fig.1 Descriptive diagram of auricle. (The origin is defined by the mid-point of 'crus-intersection' and 'incisura intertragica.')](image)

Much knowledge of external ear function has been gained through experiments on real ears and replicas and through the construction of physical models with simple geometry 4-6). Especially, (A) the auricle is functionally composed of pinna flange (helix, antihelix, lobule, etc), fossa, and concha (cybra, cavum, crus helias), (B) modal surfaces of the normal modes of the human external ear under blocked meatus conditions appear near the crus helias, between the cybra and fossa, or across the crus helias, (C) minor variations in the placement of the crus helias and the size of the channel which links the cybra and fossa have a major effect on the response frequencies and excitation of the second and third modes of the concha, and (D) the response of the external ear is strongly dependent on the "open-nose" of the concha.

We, therefore, adopt measures of describing the shape of the auricle as follows: (1) The measures are separately defined on the functional components of the auricle. (2) The coordinate system is defined by the fiducial points, line, and plane associated with the concha. (3) If the pinna flange is cut by the azimuthal plane perpendicular to the fiducial concha plane, 7 points of tangency with vertical slop or horizontal slop are defined as landmarks. (See Fig.2) The coordinates $r, \theta$ of the landmark in the section of azimuth $\phi$ are defined as the measures of describing the shape of the pinna flange. (4) The concha becomes a simple cave, if the crus helias is removed from it by the successive replacements of the tangential line in the sectional contour cut by a plane $PS^0$, which is perpendicular to the fiducial concha plane and is parallel to the fiducial concha line, and if its openings to the fossa and the eartunnel are shut. The shape of a modified concha is expressed in terms of azimuthal coordinate. The coordinate $r$ in the $(\theta, \phi)$ direction is defined as the measure of describing the shape of the modified concha. (5) The measures of the removed crus helias are defined by its attaching position to the concha, and some measures describing the shape and size of its section cut by the plane $PS^0$ mentioned above. Furthermore, the measures of the eartunnel entrance are defined by its attaching position to the concha, and some measures describing its shape and size.

MATERIALS

Replicas of the left auricles of 50 male young adult Japanese are made of dental stone plaster. Their ages range from 18 to 24. They are students at Dept. of Acoustic Design, Kyushu Institute of Design.

MEASURING METHOD

Measuring Apparatus

Two apparatuses for measuring contour of an auricle were assembled. One is used for measuring the pinna flange with the cylindrical coordinates. The other is used for measuring the concha with rectangular Cartesian coordinates. The replica is set on the apparatus with its fiducial concha plane horizontal. A measuring needle is mounted at an arm which can move in the direction of the coordinate system. Coordinates of the needle can be fed into a computer by pressing a button switch.

Measuring Method of Pinna Flange

Cylindrical coordinates $r, \theta, \phi$ of 7 landmarks in the azimuthal section of the pinna flange are mea-
sured with the azimuthal interval of 9 degrees. In the triangular fossa, its depth z in its azimuthal section is measured at the distance r from the z-axis with the distance step of 2 mm.

**Measuring Method of the Concha**

Reversal replica of the concha was made of silicone rubber and was measured. Measured are rectangular coordinates x,y,z in the sectional contour cut by the planes \( \phi \) which moves in the y direction with the interval of 2 mm, so that resultant step does not exceed 2 mm in the direction of x and z.

**RESULTS AND DISCUSSIONS**

**Statistics on Pinna Flange in Terms of Cylindrical Coordinates**

Measured data of the azimuthal section with \( \phi \) interval are so interpolated that the pinna flange shape of each subject is represented in terms of the spherical coordinates \( r, \theta \) in the section of azimuth \( \phi \) at interval of 3 degrees. Then statistics on respective \( r \) and \( \theta \) of each landmark in the section of azimuth \( \phi \) are obtained. To illustrate these statistics on the pinna flange shape with figure, let us imagine the following three auricles: (1) Average auricle, (2) Auricle which is larger than the average by the standard deviation, and (3) Auricle which is smaller than the average by the standard deviation. Fig.3 shows 50 contours of the landmark \( \phi \) and its statistics. Fig.4 shows the front view of the "average" and "larger than the average by the standard deviation" pinna flanges. Of the 7 landmarks, the standard deviation of the landmark \( \phi \), is nearly the same as the others, though it is the nearest to the origin. It may be due to the disparities in its shape.

**Fig.3** Fifty contours of the landmark \( \phi \) and its statistics.

**Fig.4** Front view of the "average" and "larger than the average by the standard deviation" pinna flanges.

**Statistics of Modified Concha in Terms of Spherical Coordinates**

After interpolating, the modified concha shape of each subject is represented in terms of the radius \( r \) in the direction with polar angle \( \theta \) and azimuth \( \phi \) at interval of 6 degrees. Fig.5 shows the statistics on the radius \( r \), using the three concha.

**Generation of Auricle Based on Statistics**

For example, a procedure of generating an auricle with average can be done in the computer as follows: (1) Set seven contours of the landmarks as frames. (2) Span surfaces between these frames by fitting elliptic or third order function in the section. (3) Hang the average shape of triangular fossa between the frames \( \phi \) and \( \phi' \). (4) Hang the average shape of the modified concha from the frame \( \phi' \). (5) Mount the average crus helias at its average position on the modified concha. (6) Make openings of the average shape and size to the ear canal and the fossa at the average attaching position.

**Fig.5** Statistics on the radius \( r \) of the modified concha.

**CONCLUSION**

A method of measuring the auricle shape three-dimensionally has been described and the statistics on the auricle shape has been obtained in 50 male young adult Japanese. The auricle shape is represented in terms of the radius \( r \) in the \( \phi \) direction in the spherical coordinate system associated with the fiducial concha plane. With the result, a model auricle based on the statistics, e.g., an average auricle, can be easily realized.

**APPENDIX**

**Definitions of Fiducial Points, Line, and Plane Associated with Concha**

1) Highest point of antihelix: the one when the side of the head is horizontally laid. 2) Crus intersection: a point on the crus helias where the crus helias intersects the projection of the crus anthelix inferior by the ray normal to the fiducial concha plane. 3) Fiducial concha liner a line passing through "crus-intersection" and "indusura intertragica". 4) Fiducial concha plane: a plane defined by the fiducial concha line and "highest point of anthelix".

**REFERENCES**

SOUND PRESSURE DISTRIBUTION IN THE EAR CANALS OF CATS

Shyam M. Khanna
Dept. of Otolaryngology, Columbia University, New York, N.Y. 10032.

Michael R. Stinson
Division of Physics, National Research Council, Ottawa, Canada K1A OOE

INTRODUCTION

Input to the auditory system is generally specified in terms of sound pressure. It was shown earlier (Khanna & Stinson, 1985) that this method of input specification is not valid at frequencies above 15 kHz, at least for those animals which showed large changes in SPL with distance along the ear canal. These changes implied a highly reflecting termination of the ear canal. Theory was presented in which the tympanic membrane and the ear canal walls were treated as acoustically hard surfaces. The theoretical calculations matched the experimental data well.

In the present paper data is presented from animals which show much smaller variations of sound pressure with distance at high frequencies. A major portion of the incident energy is being absorbed by the tympanic membrane.

1. EXPERIMENTAL PROCEDURES

The details of the surgical preparation and the measurement system have been described earlier (Khanna & Stinson, 1985). In order to improve the accuracy of positioning the probe microphone in the ear canal the sound delivery system was redesigned. The new system is shown in Figure 1. Output of an acoustic driver (D) is applied via a sound delivery tube (T) to a cut ear canal (P). The sound delivery tube is connected tightly to the ear canal via an adapter (A) and a conical ear insert (I). The ear insert handle (N) is cemented to the skull with dental cement. The position of the ear insert is thus fixed throughout the experiment. Before cementing the ear insert its position and orientation is adjusted so that its axis points to the deepest portion of the ear canal. Sound pressure is measured with a probe tube (P) attached to a ¼ inch microphone (M). The probe tube can be moved along the axis of the sound delivery tube and positioned either inside the sound delivery tube or inside the ear canal. The relative position of the probe tip is determined from the vernier scale (VS) which moves with the probe and the main scale (MS). The absolute position of the probe tip is determined from a calibration in which the vernier scale reading is recorded when the probe tip is flush with the far edge of the ear insert.

Sound pressure amplitudes and phases were measured between 0.1 kHz and 33 kHz at frequency increments of 250 Hz at each position. Probe positions were set at 0.1 mm increments over a distance of 10 mm. Silastic cast of the ear canal is taken with the insert in place after the experiment. This cast allows the determination of exact locations along the ear canal where the sound pressure measurements were made.

2. RESULTS

The variation of sound pressure amplitude as a function of distance along the ear canal is shown in figures 2a and 3a. The data shown in figure 2a is for an animal with large variations in pressure.

![Figure 2](image)

Figure 2. (a) Relative magnitude of sound pressure and (b) its phase as a function of distance along the ear canal at three frequencies.

Curves for three frequencies are shown (solid line, 30.8 kHz; dashed line, 23.3 kHz; and dotted line, 16.5 kHz). Zero on the distance scale corresponds to the deepest point in the ear canal. The probe tube position was inside the ear canal from 0 to 6 mm and inside the sound tube from 6 to 10 mm. The amplitude scale has been normalized to 0 dB. The maxima and minima are due to standing waves in the ear canal (Khanna & Stinson, 1985). At 30.8 kHz the maximum to minimum ratio is 27 dB, at 23.3 kHz the ratio is 19 dB and at 16.5 kHz the ratio is 14 dB. In figure 3a data is shown for an animal with smaller variations in pressure. The maximum to minimum ratio is 9 dB at

![Figure 3](image)

Figure 3. (a) Relative magnitude of sound pressure and (b) its phase as a function of distance along the ear canal at three frequencies.
at 29.3 kHz, 10 dB at 23.5 kHz and 8.5 dB at 16.3 kHz. It should be noted that the total frequency range over which minima can be seen over the 10 mm distance is approximately from 10 kHz to 30 kHz.

The phase of the sound pressure as a function of distance in figures 2b and 3b. Phase at the 0 mm position is used as the reference phase. In figure 2b data is shown for the same animal and frequencies as in Figure 2a. Phase remains relatively constant with distance except in a transition region where there is a sharp increase in phase of approximately 180°. Regions of this sharp phase change are centered around the position at which the amplitude curve of the same frequency shows a minimum. Figure 3b shows the phase data corresponding to the amplitude curve of Figure 3a.

3. Determination of Energy Reflection Coefficient

The energy reflection coefficient has been calculated at each frequency from the curves relating the phase of the sound pressure as a function of distance along the ear canal (Figures 2b, 3b). The theory for this method has been described earlier (Stinson, 1985). The energy reflection coefficient, \( R \), is related to the maximum slope \((\Delta \phi/\Delta x)_{max}\) of the phase curve by

\[
\frac{\Delta \phi}{\Delta x}_{max} = k \left( \frac{1 + \tan^2 \theta}{1 - \tan^2 \theta} \right)
\]

where \( k \) is the wave number \( 2\pi/\lambda \), and \( \lambda \) is the wavelength. \((\Delta \phi/\Delta x)_{max}\) was determined by using a curve fitting procedure (Stinson, 1985) using data over a 2 mm distance for which it can be assumed that the ear canal is either uniform or conical in cross section. The energy reflection coefficient is shown for two animals in Figure 4. It should be emphasized that these calculations have been made using a distance scale that corresponds to the actual probe microphones positions. A more accurate calculation will require the distance to be specified along the curved axis of the ear canal. The dotted curve in Figure 4 shows the data for an animal with high reflectance. The reflection coefficient is 0.22 at 11 kHz and increases with frequency to reach a value of .92 at 31 kHz. Thus 78% of the incident energy is absorbed at 11 kHz while only 8% is absorbed at 31 kHz. The solid curve shows data for an animal with low reflectance. The reflection coefficient is 0.1 at 10.5 kHz with increasing frequency it increases to a value of 0.38 at 18.5 kHz and then decreases to a value of 0.05 at 20 kHz. Beyond 30 kHz the reflection coefficient increases steeply reaching a value of 0.7 at 33 kHz. In this animal up to 30 kHz a major portion of the incident energy is absorbed.

Discussion

There are large variations between animals in the degree to which the incident acoustic wave is reflected by the terminal end of the ear canal. Even in the same animal the reflectance is frequency dependent. Preliminary investigations of pressure show that the degree of reflectance is correlated with the condition of the tympanic membrane and middle and inner ear. Experiments are under way to correlate the physiological condition of the ear with the degree of reflectance. The theoretical model describing the sound field in ear canals (Khanna & Stinson, 1985) cannot be applied to the case where there is significant absorption of energy. A first attempt at extending the model to include absorption at one point in the ear canal has already been made (Stinson, 1985b).

In an earlier paper (Khanna & Stinson, 1985) we were concerned primarily with animals showing high energy reflection coefficients. For such animals, because of the large variations of sound pressure with position, the measurement of sound pressure at a single position in the ear canal is not appropriate to define the acoustical input. For animals with low reflectances this problem of measurement is not so serious. However, the existence of such a large range of observed reflection coefficients raises a more fundamental difficulty. The rates at which energy is absorbed can be very different from animal to animal with nominally the same eardrum sound pressure.

In cases when the reflectance is high only a small fraction of incident energy is absorbed. Neither the incident energy nor the reflected energy describes the input because only the portion of the incident energy that is absorbed by the inner ear can be utilized to produce auditory sensation. Thus, as long as there are large inter-animal variations in reflectivity ear canal sound pressure is not a good indicator of the acoustical input. We believe that the measurement of absorbed acoustic power may prove to be a better measure to define the input.

Acknowledgement

The authors acknowledge assistance from Allen Hellard, Ravi Khanna, John Litz, David Lund, and Brian Traylor. The work of SMK was done under NIH grant 2-R01 NS 03654.

References


SOUND TRANSMISSION PROPERTIES OF THE CAT EXTERNAL EAR

A.D. Musicant, J.C.K. Chan, J.E. Hind

Dept. of Neurophysiology, University of Wisconsin Medical School, 1300 University Ave., Madison WI 53706

INTRODUCTION

Over the years, experiments in sound localization have focused most heavily upon the binaural cues of time and intensity differences between the two ears. Neither cue would seem to be able to resolve questions of front-back or up-down. Bloch (1893), Gilse and Roelofs (1930) and Musicant and Butler (1984) have shown the necessity of an undistorted pinna to make these discriminations.

Datcu (1967) theorized that direct and delayed signals provide information from one ear that could resolve the locus of free-field sound. Blauert (1971) suggested that time delays could be expressed in the frequency domain as sharp spectral minima. Shaw (1974, 1982) described the transformation of sound from the free-field to both a blocked meatus and to the eardrum of humans. For the majority of his subjects an orderly progression of minima can be described for elevations from -15° to about 60°.

Several investigators have shown that in humans the perception of elevation seems to be related to the shift in frequency of spectral minima as a sound source moves from low to high positions. Given the human psychoacoustic and physical data and the widespread use of the cat in auditory neurophysiology, we decided to measure the transformation that occurs in the cat from free-field sound to the eardrum.

METHODS

Domestic cats with clean ears had small probe tubes (1mm ID) implanted in both ears through the wall of the external canal in a ventro-posterior position. Probes were positioned flush with the inner canal wall and as close as possible to the drum (2mm) and the vertex of the canal. The ears were then arranged in a natural upright position. Thus far, in two cats, probes were coupled to 1/2" condenser microphones and, in a third, to subminiature electrets, with similar results. Stimuli were generated by applying 10us rectangular pulses to a tweeter. Microphone signals were amplified, sampled at a rate of 160 kH for 6.4 ms and averaged over 300 pulses. Speaker location was computer controlled and moved through 360° of azimuth and elevations from -36° to 90° (directly overhead). Measurements were made every 4.5° for azimuth and 9° for elevation.

RESULTS

Figs. A-F are plots of data from the left ear for three azimuthal and four elevational locations, showing transformation in dB as a function of frequency. The transformation spectra were derived by first applying an FFT to the microphone signal from the ear canal and then subtracting the spectrum obtained with the microphone in the free-field. Examination of the data indicates a number of features qualitatively similar to those described by Shaw in humans. A broad peak between 3 and 5 kH is relatively invariant in amplitude over azimuth and elevation. Between 10 and 20 kH a sharp minimum is seen in the records. This feature is present at low elevations (-18°) and shifts to higher frequencies as elevation rises. This can be seen in each panel pair. In panels A and B the first minimum appears at about
10 kHz. At 18° elevation the frequency of the min-
imum has shifted to about 13.5 kHz and at 36° the min-
imum occurs at 15 kHz. For higher elevations the min-
imum virtually disappears and a general amplitude
shift occurs across frequency. The first minimum
also shifts to higher frequencies as azimuth changes
from frontal (-27°) to more lateral (-90°) locations.
Frequencies below 5 kHz show relatively little change
with location. Frequencies above 5 kHz at -18° and
72° elevations as well as those at -90° azimuth are
lower in amplitude than at other positions. This may
indicate that the sound source is outside the area in
space directly facing the pinna.

Figures G and H are iso-transformation contours
at two frequencies. Contours were calculated from
the transformation values at each location by a com-
puter program (Precision Visuals Corp.). Phillips et
al. (1983) first described such contours for pressure
maxima. Fig. G is remarkably similar to their con-
tours at 8 kHz. A broad area is defined by the con-
tour 2 dB below the maximum. Contours in 4 dB steps
encompass even larger areas. The orientation of the
contours roughly parallels the base of the left pinna.
Fig. H is a contour plot of 12.6 kHz. Orientation of
the contours representing the maxima is similar to
that seen in Fig. G. A more complex pattern emerges
as the contours are followed to the minimum values,
especially at higher frequencies. Note that the
minima are at different loci and define narrower re-
gions than the maxima. We suggest that these spec-
tral minima, along with other features, may play a
role in localization for the cat.

Figs. I and J further illustrate this argument.
Fig. I plots values within 2 dB of the maximum at
four frequencies. Fig. J shows the same information
for the minima at the same frequencies. The contours
enclosing the maxima define broad areas. The areas
enclosed by the minima are much more sharply defined;
however, there are several loci for each minimum.
Thus the minima by themselves cannot define the unique
loci.

**ELEVATION (deg)**

![Image of Elevation Contours](image)

**SUMMARY**

We have illustrated some of the spectral trans-
formations of sound from free-space to the eardrum
of the cat. There are qualitative similarities to fea-
tures described by Shaw for the human, notably pres-
sure increases in the 3 to 5 kHz region and, more im-
portantly, spectral minima that shift in frequency as
a function of locale. These features appear to re-
sult from pinna diffraction since we found that they
are absent when the pinna is removed. We are inves-
tigating the possibility that other features, such as
depth of minima, bandwidth and the frequency of
other minima, might contribute to more specific def-
inition of loci in space. Eventually, behavioral and
physiological experiments will be necessary in order
to determine if the cat does, in fact, make use of
any of these features.

**REFERENCES**

Shaw, E.A.G. in Localization of Sound: Theory and

Supported by NIH grants NS17732 and NS07026.
INTRODUCTION: 2-PISTON MODEL

Lumped element representations of the middle ear fail at frequencies greater than 1 or 2 kHz unless special measures are taken to accommodate the mechanical complexity revealed in holographic studies of eardrum vibration (1). In the 2-piston model proposed in 1977 (2,3) and further developed in 1981 (4,5), the eardrum is divided into a small inner area overliving the malleus and a much larger outer area surrounding it. These areas are treated as rigid pistons whose vibration is a frequency-dependent mechanical impedance. Since the pistons have mechanical as well as acoustical coupling, the acoustical network representing such a system must include an ideal transformer. When Zwolinski's network (6) for the human middle ear is modified in this fashion and reasonable values are assigned to the various network elements (4), it is then possible to accommodate two significant pieces of experimental information: (a) the substantial increase in the peak vibrational amplitude in the outer eardrum, expressed as a multiple of malaror tip amplitude, that is observed at frequencies greater than 1 kHz (1) and (b) the measured values of ear canal standing wave ratio (ESWR) including recent values for the 5 to 8 kHz range (7).

EARDRUM ASYMMETRY: 3-PISTON MODEL

In the two-piston model no attempt is made to represent the lack of vibrational symmetry between the anterior and posterior zones observed by Tomndorf and Khanna. These authors found that the maximum displacement in the posterior zone was invariably greater than the maximum in the anterior zone. This is clearly shown in their contour diagram for 525 Hz reproduced here as Fig. 1a.

The present work starts with the premise that the outer area of the eardrum should be divided into two parts Sd1 and Sd2 as indicated in Fig. 1b. These are represented by rigid pistons of mechanical impedance Zmd1 and Zmd2, respectively, coupled to So by frequency dependent coupling impedances Zmd10 and Zmd20, as shown in Fig. 1c. The mutual coupling between Sd1 and Sd2 is neglected. The acoustical network representing this simplified three-piston system has two transformers as shown in Fig. 1d. The corresponding network for the complete middle ear, with representations of the middle ear cavities and ossicular chain, are similar in form to those used by Zwolinski, as shown in Fig. 2. Here the inertances Ld1 and Ld2 represent the masses associated with the areas of eardrum Sd1 and Sd2, respectively, while Cd1, Cd2, Rd1 and Rd2 represent the corresponding peripheral compliances and resistances. Similarly, C0d1, C0d2, Rd0 and Rd0 represent the coupling compliances and resistances So and Sd1, and between So and Sd2, respectively.

A reexamination of Tomndorf and Khanna's data (8) shows that the substantial increase in peak amplitude ratio, noted earlier, is marked by a transition frequency which is much lower in the posterior zone than in the anterior zone (~2 kHz vs. 4 kHz). This suggests that the minor sector in the posterior zone displaying high amplitude behaviour

---

**Fig. 1** Development of 3-Piston Model

(4,5) the upper right quadrant in Fig. 1a) is, in fact, more flexibly attached to the malleus than the complementary major sector comprising the anterior zone and the lower part of the posterior zone. Hence it seems appropriate to make Sd2 approximately twice as large as Sd1, as indicated in Fig. 1b, and to choose network values which reflect the inferred difference in flexibility (see Table 1). The mass per unit area of Sd1 and Sd2 is assumed to be 65 \( \mu \text{g/m}^2 \) and the total area of the eardrum 60 \( \text{mm}^2 \) of which 6 \( \text{mm}^2 \) is assigned to \( S_0 \). These are the values used in the fully developed 2-piston model (4,5).

The solid lines in Figs. 3 and 4 show the calculated rms velocities and phases of the three pistons at an input of 60 dB at the eardrum, the network input impedance, and the ear canal SRW as functions of frequency for the network values given in Table 1. As in earlier work, the ear canal SRW has been calculated for a cylindrical canal 7.5 mm in diameter with the eardrum perpendicular to the axis. The broken lines show the corresponding

---

**Fig. 2** Middle Ear Network with 3-Piston Eardrum

---

**Table 1. Network Values: 3-Piston Model**

| Cgs units: acoustical ohms, \( \mu \text{F, \text{mH}} \) |
|---|---|---|---|
| So=60 | Ls=23 | C=5.1 | R=250 |
| Ra=4500 | Co=0.014 | Lo=2700 | Cc=0.0035 |
| Ro=364 | Cc=0.025 | Ro=60000 | Lc=2000 |

---
THE HUMAN HEARING THRESHOLD CURVE

The solid line and symbols in Fig. 5 show various experimentally determined average pure-tone hearing-threshold levels processed and translated into sound pressure levels at the ear drum so that they can be directly compared (3). As can be seen, there are minor discrepancies at low and high frequencies but the agreement is good between 1 and 4 kHz. So, we are bound to ponder the significance of the peak in hearing threshold level at 2.6 kHz which is, of course, associated with the principal resonance of the external ear. If the sound pressure level at the eardrum required to place the median subject at hearing threshold is, indeed, greater at 2.6 kHz than at frequencies on either side, then there must be a mechanism in the middle ear or beyond which, on average, reduces the sensitivity of the ear at this particular frequency.

The peak in the hearing threshold curve is, in fact, very similar in shape and scale to the inverse of the velocity curve for the area So (i.e., the malleus response) shown in Fig. 3. This striking similarity is surely sufficient to suggest that eardrum dynamics deserves serious attention as a major determinant of middle ear transmission.

Since the characteristics of the three-piston model have not yet been properly explored, the network values presented in Table 1 are unlikely to be optimum. For now it is sufficient to note that the SWR and input impedance curves presented in Fig. 4 are not seriously in disagreement with target values based on experimental data and can probably be improved with further effort.

REFERENCES

ACOUSTIC RESPONSE OF THE SWIMBLAADER IN GOLDFISH (CARASSIUS AURATUS)

Marti Cox and Peter H. Rogers

George W. Woodruff, School of Mechanical Engineering, Georgia Institute of Technology
Atlanta, Georgia USA. 30332

INTRODUCTION

The swimbladder, the gas-filled organ used by fish to maintain neutral buoyancy, also plays a role in audition. The otolithic organs—the lagena, saccula, and utricle—in the inner ear are stimulated directly by acoustic particle motion as well as indirectly by forces due to the response of the swimbladder to acoustic pressure. Previous studies show that species with swimbladders either in close proximity to the inner ear, or directly coupled to it via the Weberian ossicles (an extension of the vertebrae) have enhanced hearing capability [1]. The mechanics of these peripheral auditory organs, however, is not understood because nothing is known about their in vivo motion in response to an acoustic disturbance. To obtain this information, a noninvasive ultrasonic system has been developed to measure their response.

MEASUREMENT SYSTEM

The ultrasonic vibration detection system is shown in Figure 1. Amplitudes as small as 25 Angstroms with a spatial resolution of 0.6 mm are measurable with this apparatus.

Figure 1. Ultrasonic Vibration Detector

The fish is placed in a controlled low frequency sound field which causes it to vibrate. It is then probed with a focused 10 MHz sound beam and echoes are detected with a second 10 MHz transducer focused on the same spot. The motion of the reflecting organ at the focus changes the acoustic path length of the ultrasonic wave. As a result of this sinusoidal variation in path length, the phase of the received signal is modulated which produces three signals at the receiver: one at 10 MHz and two sidebands at 10 MHz plus or minus the low frequency excitation.

The amplitude of the vibration can be determined absolutely from measuring the amplitude of the sidebands relative to the 10 MHz signal. The ability to determine the amplitude thus is independent of the reflectivity of the organ.

RESULTS

Several goldfish, about 4.5 cm long, both living and dead, have been examined with this device. The fish are suspended from a target post and anesthetized in a 0.005% solution of MS-222 during the procedure.
The spectra of the received echo from the in vivo anterior swimbladder before and after turning on the low frequency sound projector at 600 Hz are shown at the top and middle of Figure 2. For comparison, the response of the saccular otolith is shown at the bottom of the figure.

Notice the difference in amplitude between the 10 MHz peak and the sidebands in each case. The otolith, being much denser than the swimbladder, is moving much less. Also the swimbladder response is somewhat nonlinear as indicated by the presence of the second harmonic at 10 MHz ± 1200 Hz.

A comparison between the in vivo and the after-death responses of the swimbladder is shown in Figure 3. The amplitudes are normalized with particle velocity because measurements were made in the nearfield. The wide spread in the responses measured after death is due to the differing time spans between death and measurement of the response. The sharp rise in the after-death response past 600 Hz indicates a lowering of the resonant frequency. The bottom curve shows the change in swimbladder response of a single fish after death occurred, and again emphasizes the necessity of obtaining in vivo data.

Preliminary responses of the otolithic organs and the Weberian ossicles have also been measured. Location of these organs is verified by X-ray. Figure 4, shows the response of these organs in a single fish. These measurements were made after death; however, at the lower frequencies, the saccular otolith is moving more than the swimbladder which may indicate the gain made possible by the Weberian ossicles.

REFERENCES


Figure 3. Comparison of response of swimbladders in live and dead goldfish: in vivo (top); after-death (middle); change in response after death (bottom).

Figure 4. Response of auditory organs in a goldfish in the nearfield of an acoustic source.
A DESCRIPTION OF THE HUMAN OUTER EAR TRANSFER-FUNCTION BY ELEMENTS OF COMMUNICATION THEORY

Klaus Genuit

Institut f. Elektr. Nachrichtentechnik RWTH Aachen
Maienstr. 23, D-5100 Aachen, West Germany

1. Introduction

The description of the outer-ear-transfer-function is important for several goals and so the subject of many published papers especially by Shaw /1/. The first object is to get averaged transfer-functions which consider the typical structure as data for spatial hearing and binaural signal processing models. The second object is to simulate the directional pattern of the external ear by using a relative simple computer model in order to investigate the relationship between the transfer-function and a change of the geometrical data. This is especially useful for evaluating the influence of hearing aids. The third object is to build up an electrical external-ear-simulator for medical diagnostic use, e.g. speech intelligibility tests and improvement of hearing aids, or at last to realize spatial sound effects for audio technique. Finally a new calibratable artificial head for acoustical measurements can be build up which consists of describable geometrical parts. Such an artificial head is very important for a lot of applications like diagnosis of sound, archival registration of acoustical events, measures of headphones and so on.

The resonances, reflections, and diffractions caused by the external geometry -head, torso, shoulder, pinnnae, cavum concha, ear canal and eardrum-determine the outer-ear-transfer-function. The functional correlation between the external geometry of persons and their individual outer-ear-transfer-functions is described in /2/, /3/. There are two classes of physical effects influencing the transfer-function, the first one is depending on the direction of sound incidence, the other is not like the cavum concha- and ear canal-resonances. In the following only parameters influencing the directional pattern of the outer ear caused by reflections and diffractions are described.

2. Model to Describe the Directional Pattern

An exact mathematical description of the disturbed sound field caused by the external ear in dependence on sound incidence is unknown. But by using the KIRCHHOFF's diffraction integral /4/ it is possible to describe the influence of the transfer-function caused by head, torso, shoulder and pinnnae approximately. Only by keeping up the structure of the transfer-function and without any exact solution of the mathematical description the geometry can be reduced and the mathematical calculation can be simplified. These reductions allow an approximate estimation of the sound pressure \( p_{in}(t) \) in a point on the surface of the reference plane and result in a systemtheoretical description of a model with well known elements of communication theory, like low-, high-, bandpass-filters and time delays.

The solution of KIRCHHOFF's diffraction integrals applied to a thin disk results in four signal-paths representing the direct, reflected and two diffracted sound waves /4/. The direct and reflected part are easy to describe by the factor 1. The diffraction path is splitted off into two additional parts, one is representing the closer disk boundary and the other standing farther with respect to the sound source. So the diffraction of the disk can be interpreted with two time delays, potentiometers and a special lowpass-filter with an allpass producing a negative group-time delay (see cavum concha-reflection in the below part of fig. 1 and \( H_{in}(f) \) in fig. 2). But the time delay is never exceeding the negative group-time delay so that the system stay always causal.

For the case the disk is not closed but opened like a ring or for example shaped like the boundary of the pinnnae, then a further diffraction sound wave with phase inverted exists (see the signal paths for boundary in fig. 1, \( K_2 = -K_2, K_3 = -K_3, K_2, K_2 \) look at fig. 2). If the diameter of the opening goes to zero, the model changes to the model of the disk that means the cut-off lowpass-frequency is approaching infinity, the time-delay is approaching zero and the combination of the two potentiometers is like \( K_1 \) in fig. 2.

The approximate calculation of reflections and diffractions caused by a body like an ellipsoid, e.g. the human head, is based on the assumption that it is possible to subdivide the body into \( n \) small rings, to determine a negligible error, because the solution of the KIRCHHOFF's integrals is based on an undisturbed sound incidence. But the differences of the inner and outer diameter of the rings can be chosen very small in comparison to \( 1/4 \) wave length so that the error does not increase. The complex superposition of the \( n \) reflected and diffracted waves can be modeled...
by the transfer-function of a 6 dB/octave highpass
/1/, whose cut-off frequency is a function of the
size.

The model (labeled head in the upper part of
fig. 1) together with the parameters in fig. 2
describe the sound pressure in a point on the surface
of a body shaped like an ellipsoid in dependence on the
geometrical data and the sound incidence. This is a
general model—using only well known systems of
communication theory—which considers special cases
like the sphere and the disk.

\[ TP: \frac{H_r H_l}{H_{ps}} = \frac{1}{1 + \frac{1}{TP}} \]

\[ HP: \frac{H_{ps}}{H_{ps}} = \frac{1}{1 + \frac{1}{HP}} \]

\[ K_1 = \frac{\cos(\theta) + 0.5 \cdot r - \sin(\theta) \cdot \cos(B - \theta)}{\cos(\theta) + r - \sin(\theta)} \]

\[ K_2 = -0.5 \cdot K \left\{ (1 - r) \cdot \cos(\theta) \cdot \sin(B - \theta) \right\} \]

\[ K_3 = -0.5 \cdot K \left\{ (1 - r) \cdot (2 \cdot \cos(\theta)) \cdot \sin(B - \theta) \right\} \]

\[ f_{1, HP} = 0.5 \cdot f_{1, HP, 3} = \frac{R \cdot (1 - \cos(\theta))}{2 \cdot \pi} \]

\[ f_{1, HP} = \frac{2 \cdot \cos(\theta) \cdot \sin(B - \theta)}{R \cdot (1 - \cos(\theta))} \]

\[ T_1 = \frac{r \cdot \sin(B - \theta) \cdot \cos(\theta)}{R \cdot \cos(\theta) \cdot \sin(B - \theta)} \]

\[ T_2 = \frac{r \cdot \sin(B - \theta) \cdot \cos(\theta)}{2 \cdot R} \]

\[ K = \frac{\cos(\theta)}{\cos(\theta)} \]

\[ K' = \frac{\cos(\theta)}{\cos(\theta)} \cdot \frac{r - \sin(B - \theta)}{\cos(B - \theta) \cdot \sin(B - \theta)} \]

3. Results

When comparing the transfer-function of this
model with a measured human outer-ear-transfer-
function the directional pattern has to be weighed with
the transfer-function caused by the cavum conchae-
and earcanal-resonances. The model for these param-
eries of the outer-ear-transfer-function which are
independent of sound incidence are described in /2/, /3/, /4/. Fig. 3 shows the calculated outer-ear-trans-
fer-function of one individual ear in comparison to
the minima and maxima by six measurements at the
same ear.

4. Conclusions

It is possible to reduce the geometry of the
external ear in such a way that a mathematical estima-
tion and a systemtheoretical description of the
outer-ear-transfer-function in dependency on the
significant geometrical data of the test person lead
to similar results as a measured transfer-function.
Using averaged geometrical data the averaged
transfer-function can now be calculated. The model in
fig. 1 is build up by an electric-acoustical filter
to produce earexternal signals by headphones which are
equal to the earexternal signals by sound incidence
in freefield.

On the other hand it is now possible to calcu-
late and to construct a calibratable artificial head
for acoustic measurements which is discussed in /5/.

Literature

/7/ E.A.C. Shaw, The Elusive Connection, Localisation
/8/ K. Genuit, Analytic Description of Average
Outer-Ear-Transfer-Functions in Dependency of Sound
Incidence, 11th ICA 1983
/9/ K. Genuit, Ein Modell zur Beschreibung von Außen-
übertragungseigenschaften, Dissertation RWTH Aachen 1984
/10/ K. Genuit, Bestimmung strukturgemäßter Außen-
übertragungsfunktionen, DAGA Darmstadt 1984
/11/ K. Genuit, A Special Calibratable Artificial-
Head-Measure-System for Subjective and Objective
Judgment of Noise.

Fig. 3 Calculated outer-ear-transferfunction of one
left human ear in comparison to the minima and maxima
(--->) given bei six measurements at the same person

Fig. 2 Parameter of the elements in Fig. 1

So the complete model in fig. 1 for describing
the directional pattern of the human ear which is
turned to the sound source results in the repeated
application of the same model using the acoustical
relevant geometrical data for the determination of the
parameters in fig. 2. From each geometrical part of
the model only the dimensions \( R_x, R_y, R_z \) are the
distances from the reference point \( B \) to the
boundary, \( d \) is the thickness and their spatial
positions to the reference plane are needed. The reference
plane is the earcanal entrance \( \Phi_0(t) \). The time
delay \( T_0 \) considers in dependence on sound incidence

the different spatial positions of every parts with
respect to the reference plane. \( T_0 \) is already included
in \( T_1 \) and \( T_2 \).

The directional pattern of the other ear which is
not turned to the sound source can be described by
the same model. But the time delay \( T_0 \) is zero because
there are no significant phase differences between
the direct waves and all other waves on this side.
Using this model for binaural signal processing you
have to consider the interaural time delay.

Ironoise 86 Cambridge Massachusetts USA
OTOACOUSTIC EMISSION CHARACTERISTICS DURING MODERATE CONTINUOUS STIMULATION AND INTERMODULATION

D. T. Kemp

Institute of Laryngology and Otology, Gray’s Inn Road, London WC1X 8EE

INTRODUCTION

The acoustic characteristics of the human ear canal and ear drum include a small but significant contribution from tuned biomechanical elements within the inner ear, closely associated with the hearing process. Separation of the physiological cochlear component from the dominant pre-cochlear (ie middle ear and ear canal) part of the ear canal’s acoustic response is possible by various experimental techniques based on either the high group delay (order of 10ms) and/or the saturating nonlinearity of the former and is termed a stimulated otoacoustic emission. The existence and properties of otoacoustic emissions have influenced the development and understanding of cochlear mechanics. Specifically these support the presence of resonant elements into cochlear mechanical models. Such elements could account for the extremely low damping observed in recent direct cochlear mechanical measurements. It remains unclear whether otoacoustic emissions relate to mechanisms relevant to ‘normal’ listening levels since much of the evidence has been obtained with excitation levels near to the hearing threshold or from spontaneous emissions. This paper examines cochlear emission nonlinearities at continuous moderate stimulation levels in preparation for a detailed comparison with suprathreshold hearing acuity.

TRANSIENT EXCITATION OF OTOACOUSTIC EMISSIONS

Figure 1 shows the acoustic response of a typical human ear canal to a click stimulus. The canal is closed by the transmit/receive transducer assembly. In the low gain trace the initial high level ear canal and drum response can be seen followed after several milliseconds by a small “echo”- the acoustic emission from the cochlea. The high gain expanded trace reveals the complex structure of this delayed response component. This response grows with excitation intensity nonlinearly approximately as the square or cube root. The waveform and frequency distribution of energy in the response differs in detail from ear to ear. With transient excitation of a biological system there is the possibility that the delayed response is a ‘relaxation’ emission, not directly relevant to the response of the cochlear during normal sustained stimulation. Tone pulse stimulation shows however that the response energy is always within the stimulus bandwidth. Continuous single frequency studies have shown that the emission can be continuous at an undiminished level.

CONTINUOUSLY STIMULATED EMISSIONS

Figure 2 shows the arrangement used for the present observations. Two tones can be delivered to the ear via separate transducers embedded in a plug. A microphone registers the sound pressure in the ear canal between 500 and 2000Hz. Fr determines the frequency at which a response is to be measured. This can be f1, f2, or a combination of f1 and f2 (eg 2f1-f2), and is phase locked to f1 and f2. Amplitude and phase of the response at fr is recorded. In practice a phase locked spectrum analyser with averaging implemented in software, was used so that selected sets of combination frequencies can be examined simultaneously.

Figure 3a shows the spectrum of ear canal sound during two tone stimulation of 70dB spl, at 1920 and 2000Hz resp. A family of sidebands surround the stimulus frequencies, just as would be the case had the signals passed through a nonlinearity as in 3b. Two model nonlinear systems are illustrated in 3b. Left is a smoothly saturating one (left) which produces only the first upper and lower sidebands in any quantity. Right, is a generally linear system with a small sharply saturating component active near axis crossing. This is most like the human ear response.

The right-hand nonlinear system is analogous to the ear canal situation also for the reason that it is resolvable in to a linear and small nonlinear component. The two input stimulus tones appear directly in the ear canal response at high level...
along with small re-emitted tones at the same frequency which have passed through the nonlinearity. To examine nonlinear mechanism responsible for emissions these two components must be separated.

In the model, when the levels of f1 and f2 are made unequal (3b lower) the relative intensities of the sidebands also changes, decreasing on the lower stimulus side. In the human ear this also applies. In the example given in 3a the lower sideband exceeds the upper, although in the ear canal the stimuli are equal. This suggests that internally (at the nonlinearity) excitation at f1 exceeded that at f2, at the frequencies shown. The reverse applied at other frequencies suggesting an uneven nonlinearity with frequency.

Figure 4 shows data obtained by computation, of the relative output intensities of frequencies f1, f2, 2f1-f2 (LSB) and 2f2-f1 (USB), for varying levels of f2 (f1 fixed), on passing through a model nonlinear system. The solid line is an infinite clipping nonlinearity, the dashed is a system giving output as the cube root of input, and dotted the square root. The broad trends are the same but with sideband production greatest for a clipping type nonlinearity. When L2>LI the response at L1 and LSB is reduced, those at L2 and USB are increased. It is thus possible to determine the output of any f1 from a nonlinear system in the presence of a strong linear component, by observing the change in L1 when L2 goes from off to >L1. When L1 is similar to L2 the relative sideband intensity is strongly dependent of L2/L1.

Multicomponent swept frequency analysis

Recordings of magnitude and phase of the ear canal response to equal intensity f1 and f2 (60dB SPL) stimulation, have been made at frequencies f1, f2, and at sidebands 2f1-f2 and 2f2-f1. F1 and f2 differed by just 32Hz. Both were swept slowly through a 512Hz band at less than 10Hz/second. With this separation ear canal levels could be kept within 1/20dB of equality. Additional sweeps were made with L2 off, 10dB above LI and 20dB above, to determine the true emission level at LI, as above. Twenty four sweeps were averaged for each measurement.

Data for one ear, for the band 900 to 1500Hz is shown in figure 5. The upper solid line is the f1 level linear component, constituting the input at 60dB SPL. This has the lower phase gradient (0.5ms group delay). Data for f2 is identical except for a 32Hz shift. The lower solid line is the extracted nonlinear f1 component, with high phase gradient and 9ms group delay. Its mean level is 46dB below the f1 stimulus. The dashed line shows the lower sideband magnitude and phase, the dotted the upper sideband. Both have similar group latencies to the f1 nonlinear component from 1400ms to 1300ms, and are approximately 10dB below the f1 level. All nonlinear components show uneven frequency dependencies, with those of USB and LSB being complementary. The unevenness is roughly periodic at about 100Hz (ie 1/(group latency)). Observations in other frequency bands show similar features.

DISCUSSION

A close tone pair stimulus has been used to excite otocoustic emissions in a human ear, a moderate levels. Emission components at the stimulus frequencies and at the intermodulation frequencies 2f1-f2 and 2f2-f1 have examined. All show similar latency (9mas) but different intensity frequency functions. The use of close tone pair stimuli and sideband detection has some practical advantages over stimulus frequency emission extraction methods. Simple model calculations allow us to interpret the relative levels of these. The ear data has sidebands only 10 or 12dB below the stimulus frequency emission, which is characteristic of a cube root or sharper saturating nonlinearity. Complementary frequency dependencies of LSB and USB (amounting to a 12dB variance), could arise (see fig 4) if internally excitation at f1 and f2 differed by 4kHz (over 32Hz). The stimulus frequency emissions, but not the stimuli themselves did vary by this amount over the band, so the data may indicate cochlear irregularities at moderate levels.

REFERENCES

NONINVASIVE MEASUREMENTS AND MODELLING OF HUMAN INTRA-COCHLEAR HYDROSTATICS

R.J. Marchbanks, A.M. Martin and J.J. Tweed

Institute of Sound and Vibration Research, The University, Southampton SO17 1BJ, England

A better understanding of the hydrostatic properties of the intra-cochlear fluids, their interdependence, and relationship to cerebrospinal fluid properties, is essential to our understanding of both normal and pathologic cochlear physiology (e.g. Meniere's-like disorders, perilymphatic fistulas, cochlear window ruptures, alterbaric vertigo). Predicting the effects of hydrostatic pressure on the acoustic properties of the ear is important for the differential diagnosis of these pressure related pathologies.

The inner ear fluids maintain homeostasis by a complex variety of mechanisms including various ionic, osmotic, metabolic processes and by mechanical pressure transfer across flexible boundaries. One factor central to an understanding of these processes is the cochlear aqueduct which runs from the scala tympani into the subarachnoid space. If this aqueduct is fully patent, then the perilymphatic hydrostatic pressure will directly reflect that of the cerebrospinal fluid (CSF), and the very large pressure fluctuations of this system (Davson, 1970). Human histopathological studies suggest that the patency of the aqueduct is extremely variable between ears and, although normally patent in early stages of life, appears to become more often than not, sealed in the later stages (Wisniewski, 1978).

The scope of the present paper is limited to mechanical interactions of the perilymph and CSF. Two principle questions are addressed; firstly, does the CSF hydrostatic pressure influence that of the perilymph, and secondly, how do changes in hydrostatic pressure effect the transmission characteristics of the middle ear?

NONINVASIVE MEASUREMENTS OF PERILYMPHATIC HYDROSTATIC PRESSURE

Changes in the hydrostatic pressure of the perilymph produce a small but measurable variation in the dynamics of the ossicles and tympanic membrane (TM). This effect is primarily due to the

hydrostatic pressure exerting a force sufficient to influence the resting position of the stapes footplate in the oval window and consequently the degree of freedom of inward TM motion (Dennert et al, 1977; Brask, 1978; Marchbanks, 1980).

In current studies a movement of the TM is induced by acoustic stimulation of the stapedial reflex, and is measured with special computer-based instrumentation which will resolve displacements as small as 10⁻⁹ metres (Marchbanks, 1980, 1981, 1984; Marchbanks and Martin, 1983, 1984a).

Changing the CSF Hydrostatic Pressure by Varying Subject Posture

A statistically balanced experiment was carried out on 24 normal subjects (Age 18-30 years). The magnitude of inward-going TM volume displacement on contraction of the stapedial muscle VI, was measured in three body positions with spine angles of 30°, 60° and 90° to the vertical (Tweed et al, 1986) For each posture angle, the difference ΔVI between VI in the sitting and angled position was measured and an average made of 4 measurements regularly taken over a 20 minute period, Figure 1.

A relationship between angle and the change in VI was found which was statistically significant for all postures (1% level of P distribution). Three subjects, however, showed no such relationships.
Models of Stapes Mechanics

Marchbanks and Martin (1984b) mathematically formalised the stapedial muscle mechanics and modelled the lateral motion of the stapes with the aid of a computer. Tweed (1985) refined this model by allowing an extra degree of freedom and by independently varying the dimensions and stiffnesses of the anterior and posterior regions of the annular ligament.

The lateral stapes motion given by these models may be translated into a TM volume displacement, by first estimating the footplate volume displacement using a mean footplate area of \(3.2 \times 10^{-6} \text{ m}^2\) (Weyer and Lawrence, 1956) and a mean footplate/TM volume displacement ratio of 1:10 (Ivarsson and Pedersen, 1977; Densert et al., 1977). The models predict that for hydrostatic pressure levels of greater than \(1.5 \times 10^{-2} \text{ mm saline}\), the increase in 'inwardgoingness', ΔV1, is approximately \(2.2 \times 10^{-6}\) or \(1.8 \times 10^{-6}\) per mm saline for the Marchbanks and Tweed models respectively.

PERILYMPHATIC PRESSURE AND MIDDLE EAR TRANSMISSION CHARACTERISTICS

Changes in the hydrostatic pressure of the perilymph alter the input impedance of the cochlea by affecting the compliance of the cochlear windows. This in turn alters the middle ear transmission characteristics. The significance of this effect was predicted using the Zwislocki (1962) electrical network analogue of the ear. The stiffness/pressure relationships of the oval and round windows were derived from Ivarsson and Pedersen (1977) data, combined in series, and normalised to give at zero gauge pressure the Zwislocki cochlear-compliance equivalent, \(C_c\,\text{of} \,0.65\,\text{m}^2\text{N}^{-1}\). The middle ear transfer characteristics relating stapes velocity to input sound pressure were calculated for several different hydrostatic pressures by accordingly changing the value for \(C_c\), Figure 2.

DISCUSSION

The relationship between perilymphatic hydrostatic pressure and posture may be estimated from the current experimental results by using the stapes models to translate the difference ΔV1 into an equivalent change in pressure. This can then be compared with the expected change in CSF pressure within the cisterna magna of the cranial cavity (Davson, 1970), Figure 3. Using the model relating hydrostatic pressure to middle ear transfer function, the changes in hearing acuity with posture found by Miltich (1968) and Macrae (1972) may also be expressed as equivalent changes in hydrostatic pressure. The results of these comparisons show that changes in perilymphatic pressure reflect changes in the CSF hydrostatic pressure, as found in the cat by Carlberg (1981). Since perilymphatic pressure changes were found to occur within the first few minutes of a CSF pressure change, then it seems probable that the cochlear aqueduct was the principle mean of pressure transfer.

There were notable exceptions to this, however, in that at least 12% of the subjects tested showed no clear relationship between CSF and perilymphatic pressure changes. In this proportion of subjects the cochlear aqueduct would appear to be non-patent: a frequency of occurrence for this age group which agrees with Wloka (1978).

In general the change in perilymphatic pressure would appear to be less than that for the CSF. This could be accounted for by the network of connective tissues within the aqueduct having membraneous properties by which they are able to sustain a differential pressure between the two fluid systems. However this apparent difference could very well be reduced to zero by further refinements of the models and experimental methods.

The effect of changes in perilymphatic hydrostatic pressure on the acoustic transmission properties of the middle ear is most marked up to a frequency of nominally 800 Hz. Within this range an increase of 0.1 mm saline will cause up to an 8 dB transmission loss, Figure 3. The phase characteristics are markedly affected up to a frequency of nominally 2000 Hz with the result that an increase in pressure causes an increase in the resonant frequency of the ear.

Findings of this study are of value in an investigation of cochlear hydrops currently being undertaken. They have important diagnostic implications for cochlear hydrops and will allow a better understanding of the pathophysiology of, and chronological variations in related symptoms.

1981. UK patent 8112099
1984b. ISVR Memorandum 652.

---

**Fig 3** PERILYMPHATIC COMPARED WITH CSF PRESSURE

<table>
<thead>
<tr>
<th>Pres. mm sal</th>
<th>CSF Cistern Magna</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>X</td>
</tr>
<tr>
<td>60</td>
<td>O</td>
</tr>
<tr>
<td>120</td>
<td>X</td>
</tr>
<tr>
<td>180</td>
<td>O</td>
</tr>
</tbody>
</table>

---

**Graph**

- **X** Marchbanks 1984
- **O** Tweed 1985
- **X** Miltich 1968
- **O** Macrae 1972

- **90° 0°**
LATENCY OF TONE-BURST-EVOKED OTO-AcouSTIC Emissions

S. T. Neely and S. J. Norton

Boys Town National Institute, 555 North 30th Street, Omaha, Nebraska, USA 68131
and
Hearing and Speech Department, University of Kansas Medical Center, 39th & Rainbow, Kansas City, Kansas, USA 66103

The onset latency of transient-evoked oto-acoustic emissions (OAE) and the latency of particular frequency components within an emission are of significant theoretical interest. The dispersive nature of the transient-evoked OAE is considered one of the primary indications that the OAE is of cochlear origin. On the other hand, some researchers have expressed concern (e.g., Wit and Ritma, 1980; Rutten, 1980) that the measured OAE latency is too long to be accounted for by a model of traveling-wave reflection within the cochlea as proposed by Kemp (1980). The wave reflection model predicts that the OAE latency is twice the forward latency of an acoustic transient to the place within the cochlea which generates the reflected wave. The forward latency, however, is generally considered to be much less than half of the measured OAE latency. This apparent discrepancy has been cited as support for the existence of active elements or neural mediation in the generation of OAE.

We first demonstrate the generation of OAE in a computer simulation of cochlear mechanics and the two-way travel of transients reflected within the cochlea. We then compare our measurements of tone-burst-evoked OAE latency in human subjects with estimates of forward travel time based on recent measurements of tone-burst-evoked brainstem responses (Gorga et al., 1986) and show that the new forward travel time estimates are consistent with a reflection model for the generation of OAE.

MODEL

The transient-evoked OAE can be simulated using a linear, passive, transmission-line model of cochlear mechanics (Neely and Kim, 1986) with an impedance discontinuity at some point along the cochlear partition (CP). Model parameters were chosen to provide a frequency-place map along the CP similar to that of a human cochlea. The CP length (35 mm) was represented in the model by 400 discrete sections; middle-ear and acoustic coupler sections were also added to the model to allow stimulus presentation external to the eardrum.

A discontinuity in the mechanical impedance of the CP was created by removing all of the stiffness from two adjacent sections located about 20 mm from the basal end. This modification simulates a "lesion" about 0.2 mm wide at a place which normally represents frequencies close to 1 kHz.

The response of the CP is illustrated in Fig. 1 by successive "snapshots" of CP displacement. The stimulus was a digital representation of a 4-ms, 1-kHz Blackman-windowed tone-burst. The vertical dotted-line in Fig. 1 indicates the location of the simulated lesion. The tone-burst transient starts at the basal end of the CP in the vicinity of the stapes and reaches the discontinuity about 6 msec after stimulus onset. As the transient passes by the discontinuity, a reflected wave is generated which travels back toward the stapes. The fluid pressure at the stapes is plotted as a function of time in Fig. 2 and clearly shows that the reflected wave reaches the stapes about 12 msec after stimulus onset. The reflected wave is transmitted back through the middle-ear to the ear canal with little additional delay.

Fig. 1. Simulated displacement of the cochlear partition (CP) as a function of distance from the stapes (basal end) at four different times in response to a 1-kHz tone burst. The vertical dashed line 20 mm from the stapes indicates the location of an impedance discontinuity. The lower curve shows the CP displacement 3 msec after stimulus onset. The next curve (at 6 msec) shows the CP transient as it first encounters the discontinuity. At 9 msec the transient wave has divided into two parts: a transmitted wave continues to travel toward the apical end of the CP and a reflected wave travels back toward the stapes. In the upper curve, at 12 msec after stimulus onset, the reflected wave reaches the stapes.

Fig. 2. Simulated pressure at the stapes in response to a 1-kHz tone burst. This curve shows the time course of fluid pressure at the stapes for the same cochlear model solution illustrated in Fig. 1. The second burst between 10 and 15 msec is the reflected wave component of the cochlear partition transient. Some of this reflected wave is transmitted back through the middle ear to the ear canal (without significant additional delay) where it appears as an oto-acoustic emission.
MEASUREMENTS

We measured tone-burst-evoked oto-acoustic emissions in 7 adults with normal hearing (Norton and Neely, 1986). The stimuli were Blackman-windowed tone-bursts with frequencies ranging from 0.5 to 2 kHz, durations ranging from 4 to 8 msec, and intensities ranging from 45 to 65 dB SPL in 10-dB steps. Stimuli were presented at a rate of 16/sec to a miniature sound transducer (Eymotic Research ER-2) placed in the ear canal. The sound pressure in the ear canal was measured with a probe microphone (Knowles EA-1842). Responses to 1024 stimuli were averaged for 40 msec following stimulus onset. The averaged waveforms were processed (deconvolved) to separate the emission signal from the stimulus and the background noise. An example of a processed response waveform is shown in Fig. 3.

![Waveform](image)

Fig. 3. Measured pressure in the ear canal in response to a 4-msec, 1-kHz tone-burst at 45 dB SPL. The evoked OAE shown here has a latency of 11.4 msec.

A tone-burst-evoked OAE was observed in all 7 subjects at all the frequencies used, although emission thresholds and spectral characteristics varied significantly across subjects. Emissions from human ears tend to have more complex spectral features than the model, perhaps indicating reflected components from many different places along the CP. The emission was usually not as well localized in time as the example shown in Fig. 3 and its latency was seldom well defined. Our best results were obtained at an intensity of 45 dB SPL. Our mean estimate of OAE latency at this intensity are shown as squares in Fig. 4; the standard deviations range from 10 to 30% of the mean value. The solid line in Fig. 4 is a least-squares fitted line to the latency estimates. The latency estimates obtained by Wit and Ritsma (1980) for tone-burst-evoked OAE's ranging in frequency from 0.75 to 4 kHz are represented by the dashed line in Fig. 4 and do not differ significantly from our latency estimates for frequencies between 1 and 2 kHz.

![Graph](image)

Fig. 4. Measured tone-burst-evoked OAE latency as a function of frequency at 45 dB SPL. The squares represent the mean at each frequency of data from 7 subjects, the error bars indicate one standard deviation, and the solid line is a least-squares fitted line. The dashed line shows the AES latency estimate of Wit and Ritsma (1980) for tone-burst stimuli. The dotted line is two times the presumed mechanical component of the tone-burst-evoked ABR latency (Gorga et al., 1986).

SUMMARY

The latency of a signal through a dispersive medium such as earwax is difficult to estimate without ambiguity due to the presence of background noise, stimulus artifacts, and the recirculation of transients within the cochlea. Despite these problems, the OAE latency provides important clues about human cochlear function. After careful comparison between OAE and ABR latency data based on tone-burst stimuli we see no essential discrepancy between OAE latency, ICTT and a simple, traveling-wave reflection model for the generation of OAE's.

(Work supported by the U.S. National Institutes of Health)

REFERENCES


THE ROLE OF THE ULTRASTRUCTURE IN THE DYNAMIC RESPONSE OF THE TYPANIC MEMBRANE

Richard D. Rabbitt * and Mark H. Holmes *
+ Department of Mechanical Engineering
* Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, New York, 12180

INTRODUCTION

The tympanic membrane (TM) plays a central role in converting acoustic waves in the ear canal into mechanical motion of the ossicular chain. It is therefore important in the study of the TM to address the details of this conversion process. Of particular interest are the contributions of the geometry, ultrastructure, and the coupling of the TM to the ossicular chain and to the air in the ear canal. In connection with the first two, experimental results show that the deflection, or vibrational shape, of the mammalian eardrum is not at all "piston-like". In fact, it is highly frequency dependent, becoming quite complex at high frequencies (Khanna and Tonndorf, 1972; Tonndorf and Khanna, 1972). The vibrational shapes are directly related to the stress field within the TM ultrastructure, and so, a change in the deflection can result in a substantial change in the stress imparted to the malleus by the fibrous membrane. Because of this, in order to understand how the TM converts acoustic waves into mechanical motion, we must study both the frequency dependent vibrational shapes and the related internal stress fields.

FIBER COMPOSITE MODEL OF THE EAR DRUM

Beginning with Helmholtz (1869), several authors have studied the behavior of the eardrum by formulating distributed parameter models. These "continuum type" descriptions differ from one another primarily in the constitutive law used to describe the behavior of the TM ultrastructure. Most models neglect the fibrous structure (Frank, 1923; Funnell, 1983), while others consider the fibrous structure most important (Helmholtz, 1869; Tonndorf and Khanna, 1972). This problem can be partially resolved by modeling the actual TM ultrastructure as an ideal fiber composite material and applying strength of material methods in order to approximate its behavior.

As a first approximation the pars tensa is modeled as consisting of a set of locally orthogonal radial and circumferential fibers imbedded in a single base material as indicated in Fig. 1. Since the fibers are orthogonal, selecting the local x axis to coincide with a radial fiber results in y running approximately along a circular fiber. Based on this idealization the bending, membrane, torsional and the shear stiffnesses may be roughly estimated (Rabbitt and Holmes, 1986). These parameters relate the deformation field to the internal forces and moments fields of the TM. These relationships, for a general fiber composite shell, can be written as (Jones, 1979)

\[
\mathbf{N} = \mathbf{Q} \mathbf{e} + \mathbf{B} \mathbf{k} \\
\mathbf{M} = \mathbf{D} \mathbf{k} + \mathbf{B} \mathbf{e},
\]

where \( \mathbf{N} \) is the normal force vector, \( \mathbf{M} \) is the moment vector, \( \mathbf{Q} \) is the membrane stiffness matrix, \( \mathbf{B} \) is the membrane-bending coupling matrix, \( \mathbf{D} \) is the bending stiffness matrix, \( \mathbf{e} \) is the strain vector and \( \mathbf{k} \) is the change in curvature vector.

Although both membrane and bending stiffness are included in the above constitutive relationships, the lamina may have the effect of increasing the magnitude of the membrane stiffness matrix \( \mathbf{Q} \) relative to the bending stiffness matrix \( \mathbf{D} \). This is because the fibers are relatively stiff in comparison to the mucous and epidermal layers, causing the TM to appear much stiffer in local extension than in bending. This relatively large extensional stiffness magnifies the importance of the membrane type restoring forces and partially diminishes the importance of the bending type restoring mechanisms. Also, the internal forces and internal moments appearing within the membrane interact directly with the malleus. Of the two sets of primary fibers in the pars tensa, the radial fibers play the major role in coupling the eardrum to the malleus. This is a direct outcome of their orientation as well as their strong attachment to the manubrium (Graham, et al., 1978; Shimada and Lim, 1971).

MODEL OF THE OSSICULAR CHAIN

The stress in the radial fibers represents a majority of the force imparted on the ossicular chain by the eardrum. When studying the dynamics of the ossicular chain, an integral of this stress appears as a forcing term in the ossicular chain equations. Due to this coupling, the two models must be solved together and can not be viewed as entirely separate entities. The asymptotic solution found by Rabbitt and Holmes (1986) for the general TM is flexible enough to account for a three dimensional model of the entire middle ear network. If we model each bone as a rigid body coupled to the adjoining bones, ligaments, muscles and structures, then it is possible to derive a set of ordinary differential equations describing their coupled behavior. These equations contain the details of the geometry of each bone as well as the geometry and constitutive laws for each of the attached ligaments and muscles. In addition, the incudomeatal joint and the incudostapedial synovial joint can be included directly. Although the resulting set of equations is quite large, many of the natural frequencies are outside of the audible range and hence it is possible to nondimensionalize the system of equations and consider only the distinguished terms within the audible range. Using an approximate 6 degree of freedom system to represent the ossicular chain and its
The asymptotic solution of Rabbitt and Holmes (1986) yields the low frequency result shown in Fig. 2 for a cat.

![Diagram of ear canal and eardrum](image)

**Fig. 2.** Model results showing vibration of a cat eardrum at 600 Hz. The dark fringes represent regions of constant amplitude normal to the plane of the annular ring equally spaced on an absolute scale. The dotted line on the sine wave indicates the instant in time that is being viewed relative to the pure tone forcing signal. The dotted double arrow shows the direction of rotation of the malleus at that particular time projected onto the plane of the annular ring, and the solid double arrow indicates the direction of the resultant moment imparted on the malleus by the TM at that instant (Rabbitt and Holmes, 1986).

**DAMPING MECHANISMS**

As the frequency of the acoustical signal is increased, two types of transmission losses are experienced across the TM. The first type of loss is a result of damping within the TM ultrastructure while the second loss is due to the reflection of acoustic energy back out of the ear canal. The internal damping is a dissipative loss in that mechanical energy is converted into a change in heat and entropy. The reflected energy is not dissipative but is a loss to the hearing system and is sometimes termed radiation damping.

The internal mechanical damping is a result of the fiber composite cellular structure of the membrane. When the TM changes its local curvature, the mucous and epidermal layers are forced to change shape. Since these layers are water intensive cellular structures, the deformation of the mucous and epidermis is expected to be nearly equivoluminal. This results in a flow of the viscous fluid within the cells. In addition, the viscous interaction between, and within, the cell walls also contributes to the structural damping. Since all of these effects are a result of changes in curvature of the material, the structural damping is of 'bending type'. The bending type damping can be written as a series in the rate of change of the two principal curvatures of the drum. The first term in the series is a linear bending type structural damping.

The mechanism providing the radiation damping is quite different. When acoustic waves strike the TM the eardrum deflects, and therefore, absorbs some of the energy of the wave, while the remaining energy is reflected (both into the ear canal and into the tympanic cavity). This is a result of the coupling between the acoustic equa-

**REFERENCES**


OTOACUSTIC EMISSIONS – INTERACTIONS, CORRELATION WITH THRESHOLD MICROSTRUCTURE, AND VULNERABILITY TO ASPIRIN CONSUMPTION

Glennis R. Long, Arnold Tubis and Kenneth Jones

Department of Audiology and Speech Sciences and Department of Physics, Purdue University, West Lafayette, Indiana 47907, U.S.A.

Edward M. Burns

Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98198, U.S.A.

Many human ears generate sounds (otoacoustic emissions) in addition to transducing sounds into neural impulses. The existence of such sounds was predicted by Gold in 1948 as a consequence of hypothesized active filtering processes, but it was not systematically investigated until the late 1970’s. Kemp (1978) started systematic investigations of otoacoustic emissions in an attempt to explain the existence of consistent sharply tuned maxima and minima in human hearing thresholds (now known as threshold microstructure). Since 1978 many studies (reviewed in Zwislocki and Schloth, 1984; Zurek, 1985) have reported the following types of otoacoustic emissions:

a) Spontaneous otoacoustic emissions which can be recorded in the absence of external stimulation and in most instances can be recorded at approximately the same frequency and intensity for several years. Approximately 40 percent of human ears investigated to date (including infant ears) have been found to have spontaneous otoacoustic emissions (reviewed in Strickland, Burns and Tubis, 1985). A few of these individuals have alternate state emissions in which two different patterns of emissions appear at different times. Adding external energy near the frequency of one of these alternate state emissions evokes the emissions of that state while suppressing the emissions of the other state (Burns et al., 1984; Tubis et al., 1986).

b) Delayed evoked otoacoustic emissions which are recorded in response to transient stimulation of the ear. These emissions saturate at about 20 dB SPL and the spectra is unique to the ear being tested. The delay between stimulation and appearance of the emission is a function of the frequency and is related to the travel time of the basilar membrane (reviewed in Zurek, 1985). Spontaneous emissions are entrained by the transient stimulus and appear as evoked emissions (Wit Langevoort and Ritma, 1981).

c) Synchronous evoked otoacoustic emissions or stimulus frequency otoacoustic emissions which occur at the same frequency as an external stimulus leading to either enhancement or reduction of the sound pressure in the ear canal in the presence of a low level external tone. The amplitude of the evoked emission then can then be extracted by a vector analysis (Kemp and Chum, 1980; Zwislocki and Schloth, 1984). The existence of synchronous evoked emissions can be detected by measuring the Fourier component of the ear canal pressure at the external tone frequency (f), as f is swept through the range of interest. Spontaneous and evoked emissions appear as fluctuations of the sound pressure level and are associated with rapid changes in the phase and amplitude of the ear canal signal. If a tone is swept slowly through a frequency range near large spontaneous otoacoustic emission it can be shown (Zwislocki and Schloth, 1984) that the spontaneous emissions is synchronized by the external tone when it is sufficiently close in frequency to the emission. The range of synchronization is level dependent and can be characterized by a synchronization tuning curve. Surrounding this area is a region of partial synchronization which is unstable leading to amplitude fluctuations in the ear canal not attributable to simple beating between the external stimulus and the emission (Zwislocki and Schloth, 1984). This behavior is very similar to the transition to synchronization of a limit-cycle oscillator driven by a sinusoidal force (e.g. Machlup and Sluckin, 1980); and it supports the modeling of the source of these emissions as limit cycle oscillations in the cochlea due to some type of biochemical feedback amplification. Further evidence in support of this picture is obtained from the statistical properties of spontaneous emissions (Bialek and Wit, 1984).

d) Distortion product emissions occur when two similar frequency tones are present in the cochlea. Most studies have measured difference tones (F1-F2) or combination tones (2F1-F2) evoked by two external tones (reviewed in Zurek, 1985). Although distortion product emissions are easily recorded in new-born human mammals, they can only be measured in the human ear canal if the distortion product falls near an evoked otoacoustic emission (Wilson, 1980; Brown and Kemp, 1984). We have determined that distortion product emissions also can be recorded at an external tone and a spontaneous emission, or for two spontaneous emissions so long as the distortion product occurs near an evoked emission (Burns et al., 1984; Tubis et al., 1986).

Most attempts to correlate otoacoustic emissions with threshold microstructure (reviewed in Zwislocki and Schloth, 1984) have found that while spontaneous emissions are always associated with pronounced minima in threshold microstructure, not all threshold microstructure is associated with spontaneous emissions nor does the size of the threshold minima correlate with the size of the spontaneous emission. If delayed and synchronous evoked emissions are also considered, the correlation with threshold microstructure improves. Measures of threshold microstructure obtained from different individuals who may be using different response strategies confounds attempts to determine the impact of otoacoustic emissions on threshold microstructure. The relationship would be easier to determine if changes in threshold microstructure could be monitored in one individual while manipulating the level of the otoacoustic emissions.

Spontaneous emissions can be reduced below the noise floor (McFadden and Pfl Emmert, 1984) and delayed evoked emissions reduced (Johnson and Elberling, 1982) by aspirin consumption. By monitoring spontaneous, delayed and synchronous evoked emissions, and threshold microstructure in four subjects during intake of 3.9 g of aspirin a day for three or four days and recovery, we were able to determine the impact of changes in spontaneous and evoked emissions on threshold microstructure (see Long, Tubis and Jones, 1986 for report on two of the subjects). Spontaneous emissions disappeared first with little initial impact on evoked emissions. Evoked emissions were then reduced to near the noise floor for all but one of the subjects, who plateaued with one spontaneous emission still above the noise floor and with evoked emissions reduced but still well above the noise floor. The size of threshold microstructure, measured as the ratio of the threshold minima to the neighboring threshold maxima, closely followed the size of the synchronous evoked emissions. Reduction in threshold microstructure in the three of the subjects who lost all spontaneous emissions started by marked reduction of the threshold maxima with little change in the threshold minima. The fourth subject, who never lost all otoacoustic emissions) showed marked reductions in both threshold...
old maxima and threshold minima. It is interesting to speculate that this reduction of threshold maxima with reduction of the emissions is associated with the release from masking by the otoacoustic emissions.

Threshold microstructure is reminiscent of the dips in tone-on-tone masking curves, due to detection of beats, when the frequency of the stimulus is close to the frequency of the masker (Wegel and Lane, 1976, Vogten, 1978). In an attempt to determine if the threshold microstructure can be simply modeled as a manifestation of tone-on-tone masking, we found that although similar (but not identical) shape threshold microstructure can be induced by near-threshold external tones, the two types of microstructure were affected very differently by the use of narrow-band noise probes instead of tonal probes. The naturally occurring microstructures were not significantly reduced, while those produced by the introduction of a near-threshold tone were eliminated (due to the elimination of amplitude fluctuations as a response).

The threshold minima in threshold microstructure near otoacoustic emissions thus cannot simply be attributed to the detection of beats between the probe stimulus and the otoacoustic emission. This is not surprising in view of the previously mentioned properties of emissions which are entrained or synchronized by similar frequency external tones.

Work is now in progress to use a simple driven limit-cycle oscillator model to simulate a sound which gives a psychophysical percept similar to that of an external tone whose frequency is slowly swept through a range containing a strong spontaneous emission. The success of this endeavor would provide additional evidence for active cochlear feedback mechanisms whose underlying basis is presently unknown.

REFERENCES


FORWARD AND BACKWARD COCHLEAR WAVES

Joel Zwislocki
Institute for Sensory Research, Syracuse University, Syracuse, N.Y. 13209, U.S.A.

Normally, sound enters the cochlea through the middle ear and the oval window. It is propagated there in the form of transversal waves on the cochlear partition that runs the length of the cochlear canal. When sound reaches the cochlea through its bony walls, via bone conduction, the wave pattern remains unchanged. However, when the sound source is located directly on the partition, two waves result, as was demonstrated on an electrical analog of the cochlea. One runs forward, toward the cochlear apex, the other backward, toward the oval window. There, part of its energy is reflected, producing standing waves, and part is transmitted via the middle ear to the ear canal where it can be detected.

Whether sound is delivered to the cochlea through the middle ear or skull bones it produces the same pitch sensation. This implies that the wave pattern in the cochlea remains the same. Békésy (1952) showed that, indeed, sound reaching the cochlea through bone conduction can be cancelled out by the same sound transmitted through the middle ear. Subsequently, Békésy (1942, 1945, 1955) demonstrated on mechanical models of the cochlea, that the wave pattern on the cochlear partition did not depend on the location on the cochlear capsule through which it was transmitted. Under some conditions the waves were forced to run paradoxically toward the sound source, rather than away from it, as common sense would dictate, but they always traveled from the cochlear base toward the cochlear apex. This invariant direction of wave travel in the cochlea became a dogma.

No wonder that Kemp's (1978) discovery of acoustic cochlear emissions that could be detected in the ear canal was at first greeted with disbelief. Not only did they require a sound source in the cochlea but also propagation of cochlear waves toward rather than away from the stapes. However, a simple analysis shows that the phenomenon of Kemp's (1978, 1979) acoustic cochlear emissions and Békésy's (1965) paradoxical waves are not in conflict.

Over 30 years ago, I had the opportunity of explaining the paradoxical waves by a rather simple mathematical analysis based on Hamilton's principle of minimum energy and some other straightforward considerations (Zwislocki, 1953). Without recourse to mathematics, the process can be summarized with the help of Fig. 1A as follows:

The figure shows a schematic drawing of a straightened out cochlea. The central line indicates the cochlear partition, and the curved lines at the left end of the canal, the oval and round windows, with the latter at the bottom. The dashed lines around the perimeter of the canal indicate what happens when the cochlea is compressed by a mechanical wave traveling in the skull bones. An oscillating compression must generate a compressional wave in the cochlear fluids, which travels through the cochlea at a much greater speed than the transversal wave on the cochlear partition. The relatively high speed is produced by the large bulk modules of elasticity of the fluids. The oscillating pressure generated in the cochlea by the compressional wave is partially equalized through the cochlear windows which have a relatively low acoustic impedance. If the cochlea were perfectly symmetrical relative to the cochlear partition, the latter would not be displaced at all by the fluid motion accompanying the compressional wave and the pressure equalization through the windows. However, the cochlea is highly asymmetrical being connected to the vestibular system on one side of the partition and having the highly flexible membrane of the round window on the other. As a result, a net fluid flow must occur through the plane of the partition. This flow causes a displacement of the partition. It should be noted that the acoustic impedance of the helicotrema at the cochlear apex is high at all audible frequencies, but the lowest, and does not short-circuit the fluid flow. Although the compliance of the partition increases by a factor of over 100 from the cochlear base to the apex, this compliance has practically no effect on the displacement of the partition produced by the compressional wave because of the relatively high bulk modules of the fluids. The compressional wave acts like a velocity, or current source and deflects the partition practically uniformly along the entire cochlea, as indicated in Fig. 1A by the dashed line. A uniform displacement of the partition over its whole length means that maximum potential energy is stored in it near the windows, where its stiffness is the greatest. Once the partition is displaced, a transversal wave must be generated on it and must be propagated along the negative energy gradient—from the cochlear base toward the apex.

These considerations mean that the transversal cochlear wave generated through fluid pressure depends exclusively on the mechanical properties of the partition and on the coupling of the partition to the mass of the surrounding fluid but not at all on how and where the sound pressure is delivered.

A very different situation arises when the sound source is contained in the partition itself, as must be assumed for acoustic cochlear emissions, whether spontaneous or externally excited. Sound energy is then concentrated at the source location. If the source produces a local vibration of the partition, as indicated in Fig. 1B, the vibration is propagated as transversal waves down the negative energy gradients, one pointing toward the cochlear apex, the other toward the base. The wave traveling toward the apex behaves like a usual cochlear wave produced by stapes vibration. Its amplitude decays to zero because of dissipative losses, and no wave reflection takes place (e.g. Zwislocki, 1983; deBoer, 1983). On the other hand, the wave traveling toward the windows finds itself in a region where the impedance of the cochlear partition is dominated by stiffness, and dissipative losses are negligible. It reaches the windows without any appreciable energy loss and is reflected there because of an impedance mismatch.

Some of the possible resulting wave patterns are shown in Fig. 2. The dark band indicates the location of the source. The patterns have been obtained on an electrical transmission-line analog of the cochlea with 96 sections, constructed according to the transmission-line differential equation derived for the cochlea (Zwislocki, 1948, 1965, 1972). The cochlear model can be connected at its input to electrical networks with various impedances. The top trace shows the wave pattern that arises when the cochlear model is loaded with an impedance equal to the characteristic impedance of the transmission line. No wave reflection takes place and, as a consequence, no standing waves arise. The pattern of the second trace establishes itself when the loading impedance is infinite, and a total wave reflection takes place. A standing wave to the left of the dark band is clearly visible. Finally, the bottom trace shows the pattern that should arise under more natural conditions. The cochlear model is loaded with a network representing the mechanics of the middle ear (Zwislocki, 1965) and another network representing the
ear canal (Zwislocki, 1965). The wave pattern appears to be intermediate between those of the top and second traces. A standing wave is clearly visible, but the nodes are less pronounced than in the second trace.

Fig. 2. Wave patterns produced by "spontaneous emissions" in an electrical transmission-line model of the cochlea. Dark vertical band indicates the source location. Top trace—no wave reflection at the cochlear windows; middle trace—total reflection; bottom trace—partial reflection (see text).

Two comments should be made with respect to cochlear wave reflections exemplified in Fig. 2. First, the waves are reflected only once—at the windows. There is no reflection at the location of sound generation, unless the source produces an appreciable load on the partition. Lack of wave reflection at the sound source is clearly demonstrated by the top trace. Second, the amplitude of the partition vibration at the source location is clearly affected by the standing wave pattern. As a consequence, it depends on the wave reflection at the windows. This means that the state of the middle ear and occlusion of the ear canal must be expected to affect the acoustic cochlear emissions (Schloth and Zwicker, 1983).

Finally, Fig. 3 shows the voltage analog of sound pressure in the ear canal as a function of the frequency of the model cochlear emissions. The source was always placed at the cochlear location whose best frequency was equal to the source frequency. It produced a constant voltage of 2V at the location of its application. Note that the transfer characteristic of Fig. 3 shows a broad maximum between about .4 and 1.5 kHz. This is approximately the frequency range of most of the recordings of spontaneous cochlear emissions (Zurek, 1985, for review).

The model wave patterns of Fig. 2 indicate that the acoustic cochlear emissions must be produced by oscillators located in the cochlear partition and cannot arise from multiple wave reflections. This is in accord with several mathematical analyses which have led to the conclusion that practically no wave reflection occurs in the cochlea during wave propagation from the stapes toward the apex (e.g. Zwislocki, 1983; deBoer, 1985).

**References**


---

**Fig. 3.** Voltage analog of acoustic cochlear emissions measured in the model ear canal as a function of the frequency and location of the source. The source frequency coincided at every location with the best frequency. Ear canal was closed at a distance of 1.5 cm from the tympanic membrane.
COCHLEAR LENGTH, SPIRAL TURNS AND HEARING

C. D. West

Department of Neurology, Harvard Medical School and Beth Israel Hospital, Boston, Massachusetts 02115; W. R. Fenwal State School, Waltham, Massachusetts and Department of Anatomy, Boston University School of Medicine, Boston, Massachusetts 02118, USA.

Because of failures to find differences in the macromechanical properties of curved as opposed to straight models of the cochlea (von Bekesy, 1970; Steele and Zais, 1955), the spiral form of the cochlea is commonly thought to have little or no function in hearing. Some have described it as simply a peculiarity of mammalian development. Others have considered it an economic means of house-

The interaural distance has been shown to vary inversely with the upper limit of the audible frequency range (Masterton, Hefner and Ravizza, 1959), the relationship of the basilar membrane length and number of turns in the cochlear spiral to audible frequency range has remained undefined.

This study attempts to explore that relationship.

METHODS:

Cochlear casts were made by the paraffin/cell- 

loidin double impregnation method (Gray, 1908) (see West, 1985). The number of turns in the cochlear spiral was counted starting at the round window and following the spiral course of the cochlea up to the apex. Below the round window, the cochlea moves away from a spiral path and forms the basal hook. The basal hook was not included in the measurement of spiral turns. The number of half-turns in the cochlear spiral was figured by counting the number of inter-

sections made by the cochlear duct and an imaginary line projected from the round window through the central axis of the spiral (see Fig. 2).

FIG. 1. Cochlear cast and tracing with reference line for counting the number of spiral turns.

When this method was applied to photographs of cochlear casts prepared by Gray (1908), the results agreed with the measurements of cochlear spiral turns listed in his tables. Similar agreement was obtained by applying the method to the projection drawings of cochlear spirals made by von Bekesy (1960).

Basilar membrane (or cochlear partition) lengths were taken from published measures in the literature, and audible frequency ranges at 60 db SPL were those used by Hefner and Masterton (1980) and Hefner and Hefner (1983) (see West, 1985). The audibility curves, basilar membrane lengths and spiral turns of the cochlea of the following species were compared: chinchilla (Ch), cat (C), cow (Ol), elephant (Eh), guinea pig (GP), man (Mm), mouse (Mm), rabbit (Rb) and rat (Rt). Initials found after the common name were used to identify each species in Figs. 2-5.

RESULTS:

When species were compared across orders, no re-

lationship was evident between the number of spiral turns and the basilar membrane length; $r = 0.04$ (NS). A relationship within an order (Rodentia) also failed to meet statistical significance; $r = 0.86$ (NS). At 60 db SPL and log number of spiral turns correlated positively with the number of octaves in the audible frequency range independently of each other. Zero-order par-

tial correlation for log basilar membrane length and audible octaves was $r = 0.73$ ([P].05). For log spiral turns and audible octaves, it was $r = 0.80$ ([P].01) (see Fig. 2). Controlling for spiral turns and basilar membrane length, the partial correlations were $r = 0.83$ ([P].01) and $r = 0.87$ ([P].01) respectively.

FIG. 2. Number of spiral turns vs. audible octaves.

Basilar membrane length was more strongly corre-

related with the highest and lowest audible frequen-

cies and was inversely related to both; $r = -0.85$ ([P].01) for the high frequency limit of hearing at 60 db SPL, and $r = -0.94$ ([P].001) for the low frequency limit (see Fig. 3).

FIG. 3. Basilar membrane length vs. lowest audible frequency.

An even stronger relationship was obtained when both basilar membrane length and number of spiral turns were combined. With known values of these two cochlear dimensions, the following formulas were devised to calculate the audible frequency range of different species of ground dwelling mammals:

1) $2.42 = 0.994 \log$ (basilar membrane length/ spiral turns) = log (highest audible frequency at 60 db SPL),
2) \(1.76 - 1.66 \log (\text{basilar membrane length} \times \text{spiral turns}) = \log (\text{lowest audible frequency at 60 dB SPL})\).

The correlations with measured values of the sample species were \(r = -0.88\) (Fig. 0.01) for the highest audible frequency, and \(r = -0.96\) (Fig. 0.01) for the lowest audible frequency.

The number of audible octaves was subsequently derived from the calculated values of highest and lowest audible frequencies. The correlation with measured values of audible octaves from the sample species was \(r = 0.95\) (Fig. 0.01) (see Fig. 4).

![Graph](image1.png)

**FIG. 4.** Estimated audible octaves from calculated highest and lowest audible frequencies vs. measured number of audible octaves.

The number of spiral turns correlated best with the difference between the lowest audible frequency and the lowest frequency which produces peak-shift on the basilar membrane; \(r = 0.997\) (Fig. 0.01) (see Fig. 5).

![Graph](image2.png)

**FIG. 5.** Number of spiral turns vs. difference between the lowest peak-shift frequency and the lowest audible frequency at 60 dB SPL.

**DISCUSSION:**

Von Bekesy (1960) was able to observe shifts in the peak of the traveling wave envelope in freshly dissected cochleas of mouse, rat, guinea pig, man, cow and elephant when the bone over the uppermost turns of the cochlea was removed away. With stroboscopic illumination, he observed shifts in the peak of the traveling wave envelope as he lowered the frequency which drove the basilar membrane. Progressively lower frequencies shifted the peak progressively towards the apical termination of the basilar membrane. At a point near the apical end, the peak could shift no further, even though the basilar membrane could follow the period of lower frequencies. A given portion of the basilar membrane is maximally

stimulated by higher frequencies in the living animal than in the dissected cochlea. However, the relationships found among the species studied by von Bekesy in cochlear dissections should be found in the living state as well (Elredge, 1974).

It is clear from von Bekesy's and other studies, that species can hear frequencies below which are mechanically analyzed by the peak-shifting property of the basilar membrane. As illustrated in Fig. 5, the extent of this ability corresponds very closely with the number of turns in the cochlear spiral. How the spiral form functions to expand the range of audible frequencies above or below the peak-shift limit has not yet been determined.

Although the spiral turns of the cochlea appear to influence the number of audible octaves, other factors may also. For example, the chinchilla and the guinea pig, which have similar basilar membrane lengths and nearly the same number of audible octaves have different numbers of spiral turns. The chinchilla, which has fewer spiral turns than the guinea pig, has a larger bulla. This enlarged bulla may function to expand the number of audible octaves by decreasing tympanic membrane resistance at low frequencies (Drescher and Elredge, 1974).

Similarly, factors other than basilar membrane length must influence the highest and lowest audible frequencies since acoustic nerve neurons of animals within the same species respond to frequencies within the same upper and lower limits, despite differences in basilar membrane length (Liberman, 1982).

The formulas developed in this study are necessarily incomplete since they include only two of several possible factors. Nevertheless, they do provide a simple means of estimating the audible frequency range of ground dwelling mammals.

**REFERENCES**


CALCULATION OF TRANSVERSE INERTANCE OF THE COCHLEA

R.W. Guelke and A.E. Bunn

Department of Otorhinolaryngology and Internal Medicine, P.O. Box 63, Tygerberg 7505, South Africa.

The cochlea consists of two galleries, the scala vestibuli and the scala tympani. They are separated by the basilar membrane which increases in width towards the helicotrema, an opening which connects these two galleries. It has been shown that the action of the galleries is similar to a transmission line in which the compliant element increases along its length.

The inerance at right angles to the direction of transmission (the transverse inerance) is also important and at a particular point is responsible for a transverse resonance which is associated with the frequency discrimination of the hearing mechanism (ref. 1, 2). The transverse inerance results from the transverse motion of cochlear elements of which the cochlear fluid plays the dominant role.

The precise position of this resonance can be calculated if the volume compliance of the basilar membrane and the magnitude of the transverse inerance are known. Since the volume compliance has been measured by von Bekesy in his classical experiments and the place of resonance is also known (ref. 3 & 4) it is possible to estimate the magnitude of the transverse inerance. However, because there is some doubt concerning the precise magnitude of the volume compliance (ref. 5) it is advantageous to calculate the magnitude of the fluid transverse inerance from the geometry of the cochlear galleries.

Fig. 1 shows a section of the cochlea (ref. 6) and the method used to estimate the transverse inerance. Circles are drawn with centres at the mid-point of the basilar membrane with radii at constant intervals (in this case 0.2 mm). It is assumed that in the long-wave case the transverse flow through that segment of the circle will be proportional to the area outside the circle relative to the total area of the gallery. At, i.e. if it is the total transverse flow at the basilar membrane then the flow through the arc at radius r will be:

\[ \frac{r}{\text{rad}} = \frac{1}{\text{at}} \]

where \( r \) is an element of length along the cochlea, i.e. perpendicular to the plane of the diagram.

The energy stored between the radii \( r \) and \( r + \Delta r \), \( \Delta W \) is:

\[ \Delta W = \frac{1}{2} r \Delta r^2 \Delta Z \]

where \( \Delta Z \) is the inerance of the arc of radius \( r \), using the formula:

\[ \text{inerance} = \frac{\text{mass} \times \text{density} \times \text{area} \times \Delta r}{\text{area} \times \Delta Z} \]

The area (cross sectional to the transverse motion) can be defined as equal to the length of the arc, \( r_0 \), times \( \Delta Z \). By multiplying the element \( \Delta Z \) by a factor, \( f \), the coiled geometry of the cochlea can also be taken into account. \( L \) can then be defined as follows:

\[ L = f \frac{\rho \Delta Z}{\text{rad} \Delta Z} \]

and the energy stored between \( r \) and \( r + \Delta r \)

\[ \Delta W = \frac{\rho f \Delta Z}{\text{rad} \Delta Z} \frac{1}{2} r \Delta r^2 \Delta Z \]

This calculation can be used to estimate the fluid transverse inerance as long as the current through the arc can be estimated. Close to the basilar membrane itself an estimate of the inerance of the layer in contact with the membrane can be made as follows. If the deflection of the basilar membrane is assumed to be sinusoidal in the transverse direction and if \( y \) is the transverse coordinate the current through any element \( \Delta y \) can be described as:

\[ \text{Im} \Delta Z \cos \frac{\pi y}{\text{width}} \]

where \( \text{Im} \) is the maximum current density found at the centre of the basilar membrane.

The energy stored in an element \( \Delta y \Delta Z \) is equal to

\[ \frac{1}{2} \rho \text{Im} \Delta Z \cos^2 \frac{\pi y}{\text{width}} \Delta y \]

where \( \text{Im} \) is the width of the basilar membrane.

The total energy in the element \( \Delta Z \) is then

\[ \frac{1}{2} \rho \text{Im} \Delta Z \left( \int_0^{\text{width}} \cos^2 \frac{\pi y}{\text{width}} \text{dy} \right) = \frac{1}{2} \rho \text{Im} \Delta Z \text{width} \]

This total energy can also be calculated as

\[ \frac{1}{2} \rho \text{L} \Delta Z \left( \int_0^{\text{width}} \cos^2 \frac{\pi y}{\text{width}} \text{dy} \right) = \frac{1}{2} \rho \text{L} \Delta Z \text{width} \]

so

\[ \text{L} \Delta Z \text{ width} = \frac{1}{2} \rho \int_0^{\text{width}} \cos^2 \frac{\pi y}{\text{width}} \text{dy} \]

from which

\[ \text{L} = \frac{\rho}{8 \pi \text{width}} \]

This inerance is considered to be applicable to the zero point in the calculation of the transverse inerance. If a graph is drawn of \( L \Delta Z \) as a function of \( r \) as shown in Fig. 2 then the area indicated will be the transverse inerance for that particular point.

For the section shown in Fig. 1 the fluid contribution of the transverse inerance from the scala tympani is calculated as 1419 kg/m².

The calculation of the transverse inerance for long waves is only one aspect of the determination of the travelling wave pattern to be expected. It is also necessary to determine the pressure and flow patterns for shorter wavelengths where the lateral dimensions can no longer be regarded as small compared with the wavelength. In order to estimate this effect the transverse section of the cochlea is replaced by a sector of two concentric circles radii \( r_1 \) and \( r_2 \) where the inner circle is regarded as compliant and the outer circle as rigid.

An equivalent sector can be determined by satisfying the following conditions:

1. The cross-sectional area to be the same as the area of the gallery.
2. The long-wave transverse inerance to be the same as that calculated for the gallery.
3. The arc of the inner radius to be equal to the equivalent width of the basilar membrane i.e. equal to $8F/\pi^2$.

Fig. 3 shows the sector equivalent to the scala tympani section of fig. 1. The transverse inerance for the sector of fig. 3 can be calculated using the following expression.

$$L_t = \frac{2\pi n}{4\pi} \left\{ \frac{Z_0 (\theta r_1)}{Z_1 (\theta r_2)} - r_1^2 \left[ \frac{Z_0 (\theta r_2)}{Z_1 (\theta r_2)} \right]^2 \right\}$$

where $Z_0 (\theta r) = \ln (\theta r) + c K_n (\theta r)$

and $C = - \frac{\ln (\theta r_2)}{K_n (\theta r_2)}$

$In$ modified bessel function of the first kind, order $n$.

$Kn$ modified bessel function of the second kind, order $n$.

$\theta = \text{wavelength constant} = \frac{2\pi}{\lambda}$

$\lambda = \text{wavelength}$

If the values of $r_1$ and $r_2$ as calculated from the geometry are inserted in the formula given above the value of $L_t$ is estimated at 1169 kg/m$^2$.Lt for a wavelength of 4.4 mm and $\theta=1.43 \times 10^4$ m$^{-1}$.

REFERENCES

1. R.W. Guelke and A.E. Bunn.

2. A.E. Bunn and R.W. Guelke.


5. C.R. Steele.

NEURAL TUNING MODEL ANALYSIS OF AUDITORY FUNCTION

S.H. Park, S.K. Yoo, and S.H. Beack
Dept. of Electrical Eng., Yonsei Univ.
and Myong JI Univ., Seoul Korea 120

ABSTRACT

A neural tuning model is analyzed which is based on the phenomena for basilar membrane motion, and the responses of tuning curves. By assuming the cochlear fluid is incompressible and non-viscous, a pair of second order differential equations for basilar membrane motion is solved and the neural tuning response for hair cell excitation is analyzed by linearly combining the pressure across the basilar membrane with its displacement. An analog model consists of the six blocks: limiter, amplifier, threshold, high pass, low pass and notch filters. It has been constructed which can clarify the functional relationships of the modelled system.

1. INTRODUCTION

The human auditory function is one of the most complicated systems of the human body. It can be divided into three portions: the outer ear, the middle ear and the inner ear. Much attention has been focused on the past several years on the problem of the mechanical response of the basilar membrane in the inner ear to acoustic signal stimulation. Von Bekesy(1960) was the first to measure the physical motion of basilar membrane using light microscopy under stroboscopic illumination, and Rhode and Robles(1974) gave us more detailed and accurate data using a radioactive nössbauer source. Kiang and Moxon(1974) measured the neural tuning responses which transform mechanical motion into neural data. Through the use of their physiological data, it is possible to show the discrepancy between the mechanical and neural data, and to propose a quantitative mechanical mechanism for sharpening of the response of the basilar membrane motion. In this paper, we assume the cochlear fluid is incompressible and non-viscous and derive a pair of second-order differential equations. By using them, we can obtain a solution in agreement with the Rhode data, and if we combine the pressure across the basilar membrane with its displacement, we can approach the problem of explaining these unknown mechanical-neural transformation quantitatively. Because the equations describing the function of the ear in mathematical modelling are complex and their solutions are tediouas, we have constructed an analog model. With it we can analyze the automatic gain control and the neural tuning responses using an electrical circuit.

2. MATHEMATICAL MODEL ANALYSIS

Let \( u(x,t) \) be the fluid-velocity vector and let \( p(x,t) \) be the pressure. If the cochlear fluid is incompressible and non-viscous, this implies that the divergence of \( u \) is zero and Newton's law gives

\[ \nabla \cdot u = 0 \]  
\[ p = -\rho \frac{\partial u}{\partial t} \]  

where \( \rho \) is the density of the fluid. Eliminating \( u \) from equation (1) and (2), and taking the Laplace transform gives

\[ \nabla^2 p = 0 \]  

where the boundary conditions are

\[ (-\nabla p) = \delta \delta y = 0, \quad y = H, \]  
\[ = \delta \delta V, \quad y = 0, \]  
\[ (-\nabla p) = \delta \delta x = \delta \delta u, \quad x = L; \]  

\[ s \] denotes complex frequency. If the basilar membrane (B.M.) response is linear, the above equations can be matched by impedance relations in which is the Laplace transform transforms pressure to velocity ratio. The impedance in a traverse-bending system consists of mass, loss and stiffness.

\[ Z(x,s) = K(x)/s = R(x) + sM(x) \]  

where \( K(x) \) is the B.M. stiffness, \( R(x) \) is the loss, and \( M(x) \) is the mass of the B.M. If we use the Laplace equation and boundary conditions, a pair of second order differential equations can be obtained,

\[ \frac{d^2 p(x,s)}{dx^2} - 3p(x,s) = \frac{1}{\rho} W(x,s) \]  
\[ \frac{d^2 u(x,s)}{dx^2} - \frac{1}{\rho} \frac{d}{dx^2} W(x,s) \]

where \( H \) is the cochlear height, and \( L \) the cochlear length. If we transform the hydro-dynamic mechanical model results which are in agreement with the Rhode measurements into Kiang and Moxon neural data, the neural tuning response may be derived by linearly combining the displacement and pressure of the B.M.

\[ Q(x,s) = a(x)p(x,s) - b(x)d(x,s) \]

where \( Q(x,s) \) is the tuning response and \( p(x,s) \) is the B.M. displacement. At the characteristic frequency \( CF \), the pressure and displacement are approximately out of phase. Therefore near \( CF \), the two terms are approximately equal and opposite in magnitude. If we combine \( p \) and \( d \) so that the cancellation is not perfect below \( CF \), and require that the pressure term be a few percent greater than the displacement term, we may introduce the \( \delta \) term,

\[ O(x,s) = H(x,s)b(x,s) \]

where \( H(x,s) \) is second filter transfer function.

3. ANALOG MODEL ANALYSIS

An electronic analog model has been constructed which is composed of automatic gain control and filter parts based on the following conditions.

1) Human perceivable frequency is in the range of 50 Hz to 20 kHz.
2) The human ear can be sharply tuned at the CF.
3) If the sound pressure exceeds a certain limit, the sound amplitude is attenuated.
4) Very low intensity signals are not heard by the human ear.
5) If the intensity is very low, the signal is amplified with a gain of approximately 30.

The filter part controls the neural tuning and band pass characteristics (conditions 1,2). The band pass range of 50 Hz to 20 kHz is realized by a combination of low pass and high pass filters, and the notch filter is used to model the output to approximate the sharply tuned characteristic of the hair cell excitation. The automatic gain control part...
(conditions 3,4,5) consists of limiter, amplifier and thresholder. Fig. 1 is a block diagram of the analog model.

![Block diagram of the analog model](image)

**Fig. 1.** Block diagram of the analog model

4. RESULTS

In Fig. 2, we plot some typical results from our model where the pressure, velocity and neural tuning response are compared as a function of frequency. A is the magnitude of the B.M. velocity obtained by solving a pair of second order differential equations. The magnitude characteristic gradually increased from the steps to the helicotrema, and after reaching a maximum resonance, it very rapidly decreases. The magnitude of the B.M. velocity fits the Rhode measurements almost perfectly. The shape of the neural tuning curve is quantitatively very different from either pressure or velocity. It is more sharply tuned than any of the mechanical functions. This curve is obtained by linearly combining the pressure and displacement of the B.M. and it shows the difference between the hair cell excitation and B.M. movement. The frequency is logarithmically divided into 110 points.

![Pressure, velocity, and neural tuning response as a function of frequency](image)

**Fig. 2.** Pressure, velocity, and neural tuning response as a function of frequency

In the analog model, the amplitude is logarithmically limited in proportion to the input voltage and if the input is more than 9.2V, the output is restricted owing to saturation of the OP Amp. and there is a sudden drop in the filter characteristics at a stimulation frequency of about 4 kHz. Low frequencies below 50 Hz are rejected by the low pass filter and high frequencies beyond 20 kHz are limited by the high pass filter. Ripple occurs at about 20 kHz due to the Chebychev filter characteristic. A logarithmic scale was used to match the mathematical model analysis to the physiological data. The quality factor and resonance frequency can be changed by regulating the resistor component.

5. CONCLUSIONS

In this paper, the tuning curve response of a neural tuning model which is based on the phenomena for B.M. is analyzed. By assuming the cochlear fluid is incompressible and non-viscous, a pair of second order differential equations for B.M. motion is solved and the neural tuning response for hair cell excitation is analyzed by linearly combining the pressure across the B.M. with its displacement. A sharply tuned filter mechanism is obtained for the B.M. which is sensitive to different characteristic frequencies, and a travelling wave property appears which has maximum resonance near the stapes at high frequency and a minimum resonance is near the helicotrema. This explains the difference between the mechanical properties for the B.M. motion and the neural tuning responses. An analog model has been constructed which can clarify the functional relationships of the modelled system, and an automatic gain control has been developed in this model.

REFERENCES


RESEARCH ON THE PREPROCESSING BY A LATERAL INHIBITION

Authors: V.C. Dang, R. Carré and D. Tuffelli

Address: Institut de la Communication Parlé, ENSEG/INPG, 46, Av. Felix Viallet, 38031, GRENOBLE CEDEX, FRANCE.

INTRODUCTION

Aiming towards applications in speech recognition, scientists study research tools and processing techniques. When they use physiological or psychoacoustical data it is often in order to develop analysis methods rather that for the understanding of the phenomena themselves (Chistovich (1971)).

The human auditory system is a speech analysis and recognition system, but it is a too complex one and our knowledge about it is far from being complete. One approach for the research to test a phenomenon accepted as existing in the human being is to minimize it and test it over speech signal.

We have chosen here to test the lateral inhibition phenomenon described by Karnickay et al. (1973) and reported as successfully exploited in a recognition system by Zagoruiko et Lebedev (1974, 1985).

Recently, a detailed work on lateral inhibition has been realised by Shammi (1985).

Parallelly to these studies, numerous ear models have been proposed and some of them have been used in speech recognition (Alain et al. (1965), Blomberg et al. (1983), Dolamain (1982), Caeilen (1970)).

Our work is bounded to lateral inhibition testing on the frequency domain and we do not deal with time domain inhibition (Zagoruiko (1985), Shammi (1985)).

EXPERIMENTAL PROCEDURE

The original speech is first of all processed by Fourier Transform (256 points FFT), followed by the application of a lateral inhibition filter, which characteristics are shown in figure 1. This filter is composed of one central region and two inhibition regions. The output signal -S(i)- from this filter is the weighted sum of a group of components.

\[ S(i) = -C1 \sum e(j) + C2 \sum e(j) - C3 \sum e(j) \]

\[ j=1-8/2 \quad j=1-8/2 \quad j=1+8/2 \]

B1, B2 and B3 have been chosen equal to 1 Bark and C1 and C3 have been set to vary between 0.1 and 0.9, C2 = 1.

Figure 1. Lateral inhibition law

Figure 2a represents the spectrum for vowel /I/ under the following processings:

1: 256 points FFT
2: FFT + 1 Bark integration
3: FFT + lateral inhibition
4: cepstrum

RESULTS AND DISCUSSION

Synthetic vowel

Figure 2b permits to observe the role of 1 Bark integration before the inhibition filter (vowel /I/, curve 3).

We have also tested the effect of the 1 Bark integration when it is placed after the inhibition filter (figure 2c, curve 3).

Finally, we have also searched to obtain a first information about the role of a 3 Bark integration (gravity center theory - Chistovich (1974)).

We have remarked that:

- The lateral inhibition does not play a normalisation role with respect to formant amplitudes.
- The lateral inhibition behaves as a contrast amplifier and in particular, with the 1 Bark integration, as a small spectral variation suppressor.
- At low frequencies, fundamental frequency is detected rather than first formant (vowel /I/). For the other voices, fundamental frequency as well as first formant are detected. Harmonics presence at low frequencies is concordant with the results of Shammi (1985), Zagoruiko & Lebedev (1974), Chistovich (1970) and Evans (1982);
- The placing of the 1 Bark integration after the inhibition produces much more sharper formant peaks,
- A 3 Bark integration enables first formant detection but makes the low frequency harmonic structure to disappear. It permits to reorganize formants 2 and 3 in a gravity center perspective (Chistovich (1970)) or in an F2 perspective (Carlson & al. (1979));
- Formant position is precise with respect to cepstral data,
- An increase in C1, C3 and B1, B3 generates a decrement in small irregularities detection and renders more sharper the detected peaks;
- The first inhibition region has a more important action than the 3rd region.

Our system has been tested over noisy synthetic signal with

\[ \hat{S}(t) = S(t) + N(t) \]

where \( N(t) \) is white noise. We also have:

\[ \text{Fruit} = \text{Signal} \times \% \]

If \( X>10\% \), differences are neglectable. If \( X<10\% \), the spectral envelope is less well contrasted by all the processings (Cepstrum, FFT - 1 Bark, FFT - inhibition) but the second and third formant detection capability is better when the FFT-inhibition processing is used.

Logatones

Figure 3 represents the spectrum of /BAK/ from a male speaker processed by FFT plus 1 Bark integration plus inhibition.

A contrast increase with respect to a classic sonogram is observed. Fundamental frequency is clearly detectable and in contrast with 2nd and higher formants, the first formant is in general poorly represented. Transitions and weak sounds are highly emphasized here, since amplitude normalisation is done over each analysis frame.
Under different inhibition coefficients every signal irregularity in the time domain can be enhanced (smaller B2 and stronger inhibition). Shamma (1985) has reported this irregularities. Figure 4 shows this phenomenon for /BAIAB/.

A segmentation algorithm aiming to vowel detection has been tested (see figure 3).

CONCLUSIONS

Some results obtained on processing speech-like signals with a lateral inhibition module have been presented: on the one hand this type of processing shows a clear improvement in contrasts; on the other hand, we can hopefully expect to do segmentation work. We shall now test this inhibition phenomenon on female voices under various conditions.
CODING OF RIPPLED NOISE IN THE AUDITORY PATHWAY

J.H. ten Kate & M.F. van Bekkum

Applied Physics Department, Delft University of Technology, P.O.Box 5046, 2600 GA Delft, The Netherlands.

INTRODUCTION

How information relevant for pitch is coded in the activity of 8th nerve fibers to broad band signals, like speech and cosine noise, is still a matter of dispute. A favored pattern hidden in the spike train may carry this message towards the CNS (Bayliff and Gelb, 1981a, b). Well-known candidates for this transport are the spike rate, the interval between two successive spikes and the autocorrelation function of n-order intervals (n=0-20). Response of 8th nerve fibers to speech signals 'e' and 'a' were analyzed by Sacha and Young (1980), comparing the normalized spike rate to a localized measure of the interval rate. Peaks corresponding to (three) formant frequencies of 'e' and 'a' appeared to be only visible in the measure of the interval rate and not in the normalized rate.

Pitch perception to three component tones are shown to be most likely extracted from the intervals in the 8th nerve activity (Goldstein and Srulovicz, 1977), though the spike rate cannot be excluded completely (Goldstein, 1980).

The spike rate, however, is considered to be the information carrier for repetition pitch to click-pairs and to cosine noise in several studies (Bilsen et al.1975, Narins & Evans 1980, Yost 1982). As the dominant region for repetition pitch is at about four times the delay, the maximum at the second or at the first peak-valley difference in the delay histograms of low-frequency neurons in cochlear nucleus (cf below 2 kHz) appears to be in conflict with this interpretation. Moreover the decreasing number of discernible peaks and valleys in the delay histograms of these low-frequency units along with diminishing cf makes this concept also improbable, as its value lies below four.

A further study is prerequisite for collecting more data on rate as well as more interval histograms and autocorrelograms from 8th nerve fibers with cf below 2 kHz during cosine noise stimulation. Only one autocorrelogram of an 8th nerve fiber with cf = 2.2 kHz (Boerger & Gruber 1971) to cosine noise at a delay of 3.16 ms is present in literature. It is therefore the aim of this paper to report the temporal coding of rippled noise in the spike trains of 8th nerve fibers especially below 2 kHz by means of autocorrelograms and interval histograms.

METHODS AND MODELS:

The 8th nerve preparation in cat used follows the lines in the report by Evans (1979). Further details about the equipment, the acoustic stimulator etc. are described in an earlier study (Bilsen et al.1975).

Autocorrelograms and interval histograms have been programmed in accordance to other reports in literature using 20 successive higher order interval histograms for composing the autocorrelation. These autocorrelograms and interval histograms from spike trains are subjected to statistical treatment for the detection of the delay.

Models of Duffhues (1972) and of Weiss (1966) for the 8th nerve fiber are used to obtain artificial spike trains to (1-20 dB) modulated cosine noise and are analyzed by using autocorrelograms and the synchronization vector R at the chosen intensity level. The way in which the models are used here is to check strategies as guides for further research. The model study comprises: a) The statistical dependence of the autocorrelograms on the degree of phase locking of the neuron. b) The appearance of peaks at the time delay for summed autocorrelation for three units with different cf in one half octave band. c) The cross-correlation between two spike trains from two fibers with the same cf but different thresholds and spontaneous rates.

A relationship appears to exist between the number of discernible peaks in the autocorrelograms at the delay and the values of R. R was determined from period histograms to sine burst stimulation at different levels of the internal noise source of the fiber model. The number of discernible peaks in the autocorrelogram diminished with decreasing values of R till the last only peak vanished at values of R between .15 and .20. The value of the statistical value L-2NR appeared to be less than .48 at the same time. Reliable phase locking to sine burst is only confirmed in case L>13. (N=number of spikes). Sacha and Young (1980) show in their study the possibility to determine a synchronization vector H from the interval histogram. Moreover H and R are equivalent for sine bursts. The peak-values in correlograms and the measures H, E, L are related in this model study. It appeared possible to define a statistical measure A for the autocorrelograms by using Fourier transformation of the autocorrelation function.

RESULTS:

An example of an autocorrelogram of an 8th nerve fiber with cf=.25 kHz is presented in fig.1.

Fig.1. The autocorrelogram to cosine noise (72 dB SPL) with a modulation depth of 20 dB at a fixed delay of t=22.7 ms for the 8th nerve fiber with cf=.25 kHz. 1 indicates the delay of t=22.7 ms.

Statistically significant peaks at the delays 7.7; 13.3; 17.9 and 22.7 ms of the cosine noise were found in all corresponding autocorrelograms of this fiber. The peaks in the corresponding interval histograms however were decreasing in height along with increasing delays, till they were absent at a large time delay, here at t=22.7 ms.

Comparable results were obtained from 18 eighth nerve fibers with cf between .15 and 3.5 kHz. The peaks in the autocorrelograms at the time delays were absent for 8 units with cf larger than 2 kHz and from 2 units with cf resp. .35 and .70 kHz.

The peaks were either statistically detected or by comparing the autocorrelograms at zero delay to those with the selected delay subtracting both correlograms. The last procedure was also used in
case of interval histograms.

There were a few low-frequency units (also the fiber of fig.1) in which no modulation could be found in the delay histograms at arbitrary levels. The reason for it is not clear yet. Most fibers showed well modulated delay histograms like that in fig.2 with a CF of 1.7 kHz. Fibers with CF of 1.6-2.0 kHz have at least 4 discernible peaks and valleys to 20 dB modulated cosine noise.

Autocorrelograms showed several peaks at multiple values of the reciprocate of the CF (fig.3), for the same fiber as the one with 1.7 kHz, in which the peak is hidden at 4.8 ms. Similar multiple peaks are observed in the interval histograms from fig.3.

The occurrence of several peaks in the autocorrelograms appeared together with several peaks and valleys in the delay histograms at appropriate levels in our data.

DISCUSSION AND CONCLUSIONS:

One clear peak or multiple peaks in the autocorrelograms and in the interval histograms were ascribed, on basis of the model study, to fibers with a high degree of phase locking.

The absence of a peak at the time delay \( \tau \) in the autocorrelograms for 8 units with CF above 2 kHz can be interpreted by the decreasing synchronization vector with increasing CF (Johnson 1980, Evans 1980), having a value equal to zero at about 6 kHz. Some 8th nerve fibers in the low frequency region below 1 kHz are shown with maximum values for the normalized synchronization vector about .5. There is a considerable spread in the maximum values of this synchronization vector in this area, between .5 and .9. Both studies reveal, that the value of this vector may vary between 0 and .9 dependent of the intensity levels of the tone bursts. The absence of a peak in the autocorrelogram for two low frequency fibers (CF of .35 and .70 kHz) can be understood by the variation of these values for the synchronization vector between 0 and .9.

Boerger (1974) found some neurons in the cochlear nucleus (6 out of a sample of 49), which have narrow peaked PST histograms and multiple peaks in their autocorrelograms to cosine noise. He called these phase locked neurons "Time-Delay-Units. Electrical sinusoidal stimulation of the 8th nerve fiber by a cochlear implant however drive all neurons in the anterior ventral cochlear nucleus into phase-locked responses. It is probable therefore that the TDU of Boerger (1974) are caused by the number of phase locked fibers in the 8th nerve.

Another interpretation of the existence of the dominant region for repetition pitch is possible, when is considered that the time delay peak is drowned in the peaks of the cathode ray oscilloscope. The detection of a time-delay peak by the CNS will have a higher probability for delays 3-4 times the reciprocal of CF than at shorter delays, as can be seen in the interval histograms.

REFERENCES:


MODELING AUDITORY NERVE RESPONSE DUE TO A STRAIN-ACTIVATED TRANSDUCTION MECHANISM

J. Bell and M. H. Holmes

SUNY at Buffalo, Buffalo, NY 14214 and Rensselaer Polytechnic Institute, Troy, NY 12180

INTRODUCTION

In a previous study [1] we developed a model of hair cell (HC) transduction based on Corey and Hudspeth's [2] measurements of the changes of the receptor potential, individual hair bundle displacements. It was assumed in formulating the model that the receptor potential changes they observed are determined by the currents across the apical membrane of the HC and that the active area of this membrane overlies the distal tips of the stereocilia. The distinctive features of the model are that the transduction conductance depends on the angular displacement \( \theta(t) \) of the bundle, and the model incorporates a ciliary-membrane strain mechanism for the transduction conductance based on the identification of strain-activated channels [3] and the presence of cross-links connecting the cilia [4]. When the bundle is displaced the (inextensible) cross-links cause the strain \( \varepsilon = \varepsilon(t) \) in the receptor membrane to change. This alters the transduction conductance since its activation parameter \( n \) is a function of \( \varepsilon \). It is important to note that the structural components of the bundle play a significant role in our model development and the model agrees well with the direct experimental measurements of Corey and Hudspeth.

Our purpose here is to present extensions to this basic model which include: (i) individual hair bundle motion and (ii) coupling of a synaptic model which includes a mechanism for the auditory response. One of the key points in these extensions is the role intracellular calcium. In our model, we have found experimentally that Ca plays a fundamental role in the release of neurotransmitter in the synapse. It is also part of the basal membrane currents.

EXTENSION OF THE MODEL: BASAL MEMBRANE CURRENTS

We consider the HC's axoplasm as isopotential and base our model of the basal membrane on the experimental work of Lewis and Hudspeth [5]. The basal membrane current is modeled as a steady driving capacity current in parallel with an Ohmic leakage current and three specific ionic currents: one inward Ca current and two outward K currents. The Ca conductance has a voltage dependent activation. The K currents include a voltage dependent conductance and the other has a Ca dependent conductance. If we let \( n, q_h, h \) be, respectively, Ca and K activation, K inactivation, and intracellular Ca concentration, then the model takes the form

\[
\frac{dc}{dt} = k_3 g_{Ca}(m)(v-v_{Ca}) - k_5 c.
\]

(2)

Except for this last equation, and our earlier transduction current which is strain dependent, this model is of classical Hodgkin-Huxley type. Ca is assumed, during potential changes, to accumulate just inside the basal membrane. As seen in eq. (2), the rate of accumulation is assumed proportional to the Ca current flowing into the cell and it is removed by a first-order process. In eq. (1) we have invoked a quasi-steady state assumption to obtain the Ca dependent K conductance, which is given by the \( g_p \) term where \( k \) is a dissociation constant.

EXTENSION OF THE MODEL: SYNAPSE

We now couple the above model with the synaptic. In particular, a model is described which gives the times at which the postsynaptic membrane of an afferent nerve fiber is depolarized and discharges an action potential. In the simplest form of the model the events leading to the generation of a spike potential are as follows: 1) accumulation of Ca in the basal end of the HC, which is mediated by receptor potential depolarizations, increases the release of packets of neurotransmitter into the synaptic cleft; 2) each transmitter corresponds with a receptor site on the postsynaptic membrane; 3) these transmitter-receptor complexes initiate excitatory postsynaptic potentials which, after they accumulate to a certain level, cause the nerve to fire. Concurrent with the firing is the dissociation of the transmitter and receptor sites. We view this dissociation as producing nonactive transmitter units plus nonreactive receptor sites. The latter are assumed to have an independent time scale for recovery. The nonactive transmitter units, on the other hand, must be reabsorbed into vesicles in the presynaptic (HC) membrane before they can be reused. If \( R, T \), \( R^* \), \( T^* \) represent, respectively, the density of normal and nonactive transmitter units, normal and nonreactive receptor sites, and transmitter-receptor complex then the kinetics for our firing rate model take the form

\[
R + T \rightarrow C, \quad R^* + R, \quad R^* + T \rightarrow T^*.
\]

(3)

Here \( \alpha \) and \( \beta \) are constants for the rate of receptor site recovery and complex formation, respectively, and \( \gamma \) is the presynaptic reaction rate function. The latter depends on the intracellular Ca concentration \( c \), and so, this model couples naturally with eqs. (1) and (2).

The spike generation times are determined by when the density of transmitter-receptor complex reaches a threshold value \( C_{th} \). So, if \( T(t) = C_{th} \) then ' \( T(t) = 0 \) and \( R^*, T^* \) are released in the process. The kinetic model described here, including this reset mechanism, is a form of an integrate-and-fire model.

DISCUSSION

It remains to find the material coefficients appearing in the model. As suggested by Lewis and Hudspeth [5] the voltage dependent K and Ca currents in eq. (1) appear to be Hodgkin-Huxley like. In fact, it is assumed here that if Ca is constant then the K current is linear in voltage (so, \( g_k \) is constant). As in [1] we also assume that the relaxation time for the strain activated channels is very short, so \( y \) is in a quasi-steady state. With these assumptions eq. (1) reduces to
\[ \frac{dc}{dt} = \gamma(c) (S_1 - C - C) - \alpha(C - C_0) - \beta(t) - \gamma(c) \]

The functions \( H(v) \), \( n(c) \), along with \( c_m \), \( g_s \), and \( g_k \), are given in [1] and the function \( m(v) \) is determined from the experiments in [2].

In addition to the last three equations we also have the firing rate model which, from (3), is

\[ \frac{dc}{dt} = \alpha(t) (S_0 - R^* e^{-\beta t} - C) \]

and

\[ \frac{d\gamma}{dt} = \frac{\gamma(c) (S_1 - C - C) - \alpha(C - C_0) - \beta(t) - \gamma(c)}{1 + \gamma(c)} \]

where \( S_0 \), \( S_1 \) are constants representing, respectively, maximum available receptor sites and total transmitter available. According to Kandel and Schwartz [6] there are about \( 10^6 \) receptors/\( \mu m^2 \) available at an active postsynaptic area of 1-2 \( \mu m^2 \), so we have taken \( S_0 = 10^6 \). In our preliminary calculations we have set \( S_1 = 106 \) and \( C_{th} = 0.15 \). The rate parameters are more difficult to determine because of the limited experimental measurements that are presently available. To determine \( \gamma(c) \) it is known that the rate increases with \( c \), it is zero when \( c = 0 \), and it saturates as \( c \rightarrow +\infty \). To incorporate this into the model we have taken

\[ \gamma = \gamma_0 + \frac{\gamma_1 (c - C_0)}{1 + \gamma_0 c} \]

where \( C_0 \) is the equilibrium Ca concentration when there is no external stimulus. Thus, \( \gamma_0 \) represents the spontaneous generation rate of \( T \).

As explained earlier, there is a reset mechanism in our firing rate model which corresponds to the discharge of an action potential. To illustrate this suppose \( C(0) = 0 \), \( R^*(0) = 0 \) and let \( t_1 \) be the first time that \( C \) reaches threshold \( (C_{th}) \). In eqs. (4) we then have \( R_0 = 0 \). At \( t = t_1 \) an action potential is discharged and the concentrations reset. Specifically, \( C(t_1) = 0 \), \( T(t_1) = T(t_1^-) \) and \( R_0 = 0 \). For \( t > t_1 \) the concentrations again follow eqs. (4) until the next time \( t_2 \), that \( C(t_2) = C_{th} \). At this time another action potential is discharged and the concentrations reset as before except now

\[ R_0 = C_{th} e^{\beta t} \]

This process continues, producing a sequence \( t_1, t_2, t_3, \ldots \) that gives the times at which an action potential is discharged. This sequence represents the output of the model.

There are a number of properties of the auditory-nerve discharge pattern that the model should be able to address, such as the spontaneous firing rate, adaptation, saturation, etc. We have been able to show that in the case of no external stimu-

Acknowledgements: JR acknowledges support from the NSF, grant DMS 8502307, and NIH acknowledges support from the ARQ, grant DAAG29-83-K-0092.

REFERENCES

If it transpired that both signal and noise pdfs were Gaussian, the (signal + noise) pdf is also Gaussian. The transmitted information is then:

\[ I_t = \frac{1}{2} \ln \left( 1 + \frac{\sigma_S^2}{\sigma_N^2} \right) \]  

where \( \sigma_S^2 \) and \( \sigma_N^2 \) are stimulus and noise variances, respectively. Signal variance may originate from several sources, for example unsteadiness in voltage at the sound generating equipment itself. A so-called "steady" tone is, at the quantum level, continuously fluctuating and there is recent theoretical evidence that the stereocilia can be influenced by forces at the quantum-noise limit (Bialek and Schwanzler, 1988). A further source of variation is the stochastic fluctuation in the number of rapidly opening and closing transduction channels at the tip of the stereocilium (Hudspeth, 1986) that are open at any instant of time.

### Perception of a Steady Tone

A sensory receptor must sample the stimulus at discrete intervals. Experimentally the typical stimulus used is a steady tone. Perception of the continuity of the stimulus is a consequence of repeated sampling of the stimulus pdf. The receptor will gain progressive confidence in its estimate of the mean value (stimulus intensity) as the sample size \( m \) increases. According to the Central Limit Theorem, when \( m \) is sufficiently large the mean of the \( m \) values follows a normal distribution with mean value equal to the mean of the stimulus pdf and variance equal to the stimulus variance divided by the sample size, \( m \). The total entropy change (decrease in uncertainty) is:

\[ H = \frac{1}{2} \ln \left( 1 + \frac{\sigma_S^2}{m \sigma_N^2} \right) \]  

Equation (6) is independent of the exact nature of the stimulus pdf, whether it be Gaussian, Poisson or otherwise. As time passes, \( m \) increases monotonically, uncertainty becomes increasingly smaller, and receptor response to the tone diminishes. Empirically the mean and variance of a stimulus pdf can be approximated by:

\[ \sigma_S^2 = \text{constant} \cdot \mu^0 \]  

where \( \mu \) is the intensity (acoustic pressure) of the tone.

### The Fundamental Assumption

To link the concept of uncertainty with the measurable world, the fundamental assumption is made that:

\[ F = kH \]  

where \( F \) corresponds to a "perceptual variable". This variable may be taken to be the impulse rate in the primary afferent, or the sensory feeling of audition, which appears to work in parallel. For a constant number of samples, \( m \) (i.e., constant time of stimulus duration)

\[ F = 4k \ln \left( 1 + A \mu^0 \right) \]  

For large values of loudness, the second term in brackets becomes much larger than the first, and

\[ F = B + C \log \mu \]  

(the Weber-Fechner relation). In the opposing limit, expanding (9) in a Taylor series and taking only the first term produces Stevens' power law:
\[ F = D p^n \] (11)

Results

Controversy has existed for some time as to whether different methods of magnitude judgement, such as "magnitude estimation", "magnitude production" and "category rating" are linearly related. Hellman and Zwolinski (1968) gathered both magnitude estimation and magnitude production results: all show the same "entropic" shape regardless of method (see fig. 1 for entropic and power law curvefits to this data). The entropy equation gives not only a snugger fit, but a larger "Stevens index" \( n \) than the power law. Firing rates of single eighth nerve units of cats exposed to 40 ms pure tones (Wiederhold, 1970), normalised after subtracting spontaneous firing rates (fig 2), follow the same trend. Luce and Mo (1965) recorded subjects' responses to a wide range of loudness: one such set of responses is plotted in fig. 3 and again conforms well to the entropic form. Further results will be presented.

References

Hudspeth, A. J. Science 220, 1985, 745-752
Mackay, D.M. Science 139, 1963, 1213-1216
---. Percept. Psychophys. 29, 1981, 409-422

![Graph 1: Loudness (subjective) vs. Sound Pressure (log scales) after fig. 1 of Hellman and Zwolinski 1968](image1.png)

![Graph 2: Normalised Response (spikes/sec) vs. Sound Pressure (linear scales) after fig. 2 of Wiederhold 1970](image2.png)

![Graph 3: 1000 Hz tone (subjective vs. log scales) after fig. 3 of Luce and Mo 1965](image3.png)
METHODS OF MEASURING HEARING PROTECTOR ATTENUATION

E. H. Berger
E-A-R Division, Cabot Corporation
791 Zionsville Road, Indianapolis, IN 46268-1650

There exists a vast body of literature pertaining to methods of measuring the attenuation of hearing protection devices (HPDs). In a recent review, Berger reviewed and analyzed these articles with respect to the accuracy, practicality, and applicability of the available techniques. Selected highlights of his review, as well as the conclusions, will be summarized herein. For references the reader is directed to the source document.

The noise reduction provided by an HPD is popularly referred to as its attenuation, a term which lacks the precision sometimes required when describing experimental data. This problem is resolved by the use of the expressions, insertion loss (IL is the difference in SPLs at a point in the ear, with and without the HPD in place) and noise reduction (NR is the difference between SPLs inside and outside the HPD). IL readings must be taken sequentially whereas the two measurements required for an NR evaluation may be taken simultaneously. However, IL provides the most accurate assessment of hearing in SPLs relative to the unaided ear of the listener's normal when the HPD is in use. IL is approximately equal to NR plus the transfer function of the open ear.

SUBJECTIVE METHODS
HPD attenuation evaluation methods may be broadly categorized into subjective and objective procedures (Table 1). Subjective approaches rely on judgments of human subjects to auditory stimuli presented under protected and unprotected listening conditions. The difference in their responses is a measure of the HPD's IL. These methods allow evaluation of all the possible acoustic transmission paths to the unprotected ear.

The REAT paradigm consists of measuring subjects' binaural thresholds of hearing both with and without HPDs. The difference in the thresholds, the threshold shift, is a measure of the HPD's IL. REAT tests may be conducted with normal listeners in sound fields (1a) or under earphones (1b) or using hearing impaired listeners (1c).

The principal utility of subjective above-threshold procedures (ATPs) is to verify REAT data by allowing examination of: a) potential level-dependent effects and b) errors arising from masking due to physiological noise. They can also simplify measurement of attenuation under field conditions due to less stringent requirements on the ambient noise in the test chamber.

ATPs may consist of measuring the effects of wearing HPDs upon the masking of either bone-conducted or air-conducted stimuli, or the effect of HPDs upon speech intelligibility. Additional psychological techniques (2b) such as cross-modality loudness scaling, magnitude estimation, and reaction time have also been implemented. All of these methods except (2d) and (2e) can provide quantitative estimates of attenuation.

OBJECTIVE METHODS
Objective tests are also based upon the difference between two measurements, but the data result from instrument readings with transducers located in or on artificial or real heads. Either IL or NR may be measured. These tests cannot directly account for all of the sound paths to the unaided ear although it is possible to apply mathematical corrections to the measured values. Objective methods offer the potential to measure attenuation at a variety of SPLs and the acoustical test fixture (ATF) and microphone in real ear approaches can significantly expedite data acquisition as well. The first three objective methods (see Table 1) are self-explanatory. The fourth, consists of contralaterally measuring with high-level acoustical signals (+120 dB in the occluded case) the shift in aural-reflex threshold in a protected and unprotected condition for the subject's ear.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Methods of measuring hearing protector attenuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. SUBJECTIVE METHODS</td>
<td></td>
</tr>
<tr>
<td>1. Real-ear attenuation at threshold (REAT)</td>
<td></td>
</tr>
<tr>
<td>a. Sound field REAT</td>
<td></td>
</tr>
<tr>
<td>b. Headphone REAT</td>
<td></td>
</tr>
<tr>
<td>c. REAT with hearing impaired subjects</td>
<td></td>
</tr>
<tr>
<td>2. Above-Threshold Procedures</td>
<td></td>
</tr>
<tr>
<td>a. Masking</td>
<td></td>
</tr>
<tr>
<td>b. Loudness balance</td>
<td></td>
</tr>
<tr>
<td>c. Midline localization</td>
<td></td>
</tr>
<tr>
<td>d. Temporary threshold shift</td>
<td></td>
</tr>
<tr>
<td>e. Speech intelligibility</td>
<td></td>
</tr>
<tr>
<td>f. Miscellaneous psychophysical methods</td>
<td></td>
</tr>
<tr>
<td>B. OBJECTIVE METHODS</td>
<td></td>
</tr>
<tr>
<td>1. Acoustical test fixture (artificial head/ear)</td>
<td></td>
</tr>
<tr>
<td>2. Microphone in real ear</td>
<td></td>
</tr>
<tr>
<td>3. Microphone in cadaver ear</td>
<td></td>
</tr>
<tr>
<td>4. Aural-reflex threshold shift</td>
<td></td>
</tr>
</tbody>
</table>

DISCUSSION

For routine evaluation of intentionally linear HPDs (those not containing valves, orifices, or active electronics) the standardized diffuse-field REAT tests are the most appropriate and accurate approaches, and the ones for which the greatest amount of existing data are available. The preponderance of the reported research demonstrates that for such HPDs, attenuation is independent of SPL and can therefore be accurately evaluated at near threshold. The only demonstrable inaccuracy of REAT is spuriously elevated attenuation that results for certain devices due to masking of the occluded thresholds by physiological noise (Fig. 1). The errors are limited to frequencies 0.5-2.0 kHz and are typically less than or equal to a few dB, except below 125 Hz.

Other problems that have been attributed to the REAT paradigm are the need for a very quiet test room (Fig. 2), expensive and complicated instrumentation, multiple subjects, and extended test times. The low ambient noise levels required of the test chamber are unique to threshold testing, but the remaining difficulties are equally encountered with any of the other subjective procedures. In fact, multiple subjects are always necessary when information is required about how a device will fit or protect the general population.

For any of the methods the penalty for decreasing the number of subjects (or using only one particular ATF or instrumented real ear) is the...
same—the data will be less easily and accurately
generalized to the target population.

All the subjective methods as well as the
microphone in real-ear and acoustic reflex
objective methods rely upon the use of human
subjects. Since the subject/protector interface is
perhaps the single most important parameter
affecting an attenuation measurement it must be
carefully attended to during such tests. Far from
being a liability, this is one of the strong assets
of these tests, as long as the number being sought
is a measure of the protection that the device can
offer in practice when reasonably fitted and worn.
But, one must use such data with care and awareness
of the fitting techniques, and subject selection
and instruction in order to utilize them as well as
the magnitude of both intra- (Fig. 3A&B)
and interlaboratory variability.

Whether or not the response of intentionally
linear HPDs to gunfire and cannon impulses, blasts,
and explosions, is accurately represented by an
REAT test is open to question. Some authors have
reported equivalent attenuation for low-level
steady-state signals and high-level impulses,
whereas others have reported either decreasing
attenuation, non-monotonically changing
attenuation, or increasing attenuation as sounds
become impulsive and increase in level.

For the evaluation of nonlinear HPDs high
sound level test stimuli must be utilized. The
subjective ATPs such as masking, loudness balance,
and midline lateralization, will generally only be
suitable for HPDs exhibiting level-dependent
behavior at SPLs as low as 100-110 dB, since
generation of higher levels via these methods is
impractical. The test procedures are expensive,
but they are procedurally difficult, time
consuming, and offer limited resolution. Better
methods for high-level tests are the use of ATEs or
microphones in real ears, both of which allow the
use of almost any test stimuli. Cadavers have also
been used for this purpose, but procedural and
experimental complexities limit their application.

Field evaluation of HPD attenuation can best
be accomplished via "natural" methods which
facilitate the use of actual HPD wearers as
subjects. The methods most suitable for this are
the headphone-REAT approach for evaluating earplugs
and microphone in real-ear testing for earmuffs.
Regardless of the chosen technique, or even if one
measures actual effectiveness using TTS, the
possibility exists that the experimenters, their
equipment, and their testing procedures, will
influence the attitude of the employee being tested
and hence use of their HPDs. To avoid these
temporary artifacts and provide a more global
picture of HPD effectiveness, one can instead
analyze the company’s data base of annual
audiograms.

CONCLUSIONS

It is clear that no one method of measuring
HPD attenuation is best for all conditions.
However, certain procedures are either too
difficult to pursue or are so subject to
experimental artifact that they can be removed from
consideration. What remains, is a group of proven
and reliable experimental techniques that may be
employed for routine evaluations and future
research endeavors. They include sound-field and
headphone-REAT paradigms, the masking of
bone-conducted stimuli, midline lateralization,
temporary threshold shift evaluation, and the use of
acoustical test fixtures and microphones in real
ears.
VARIABILITY AND ACCURACY OF SOUND ATTENUATION MEASUREMENTS ON HEARING PROTECTORS

K. Brinkmann, U. Richter

Physikalisch-Technische Bundesanstalt (PTB), Bundesallee 100, 3300 Braunschweig, F.R. of Germany

1. INTRODUCTION

The present standard procedure for the determination of the sound attenuation of hearing protectors is specified in the International Standard ISO 4869.1. Radically it depends upon determinations of the threshold of hearing of a number of subjects with and without the hearing protector being worn. The tests are performed in a diffuse sound field using a number of third-octave bands of pink noise in the range from 63 Hz to 8000 Hz.

Though this procedure has been used in principle in many laboratories for many years, only very little is known about the measurement uncertainties involved and, moreover, the general validity of the results with respect to the application of hearing protectors at high sound pressure levels is questioned from time to time. The present study extends the results of an earlier paper /2/ towards a quantitative evaluation of both the variability and the accuracy of hearing protector measurements.

2. EXPERIMENTS

Two experiments have been performed:

Experiment A consisted of 3 independent measurements of the sound attenuation of two different hearing protectors of the ear-muff type (Wilson 35B and Silenta Super) on different days with a group of 30 test subjects using the test procedure specified in /1/. One of the protectors (Silenta) has muffs with foam-filled cushions, the cushions of the other protector are liquid-filled. Both protectors use metal headbands and can be considered to represent good-quality protectors which can be easily fitted to most subjects. Their headbands were applied at forces of 6.9 N (Wilson) and 9.6 N (Silenta) for average head sizes. The variation of these values with varying head dimensions is rather small.

The test subjects compiled with the requirements of /1/ regarding hearing threshold levels and the repeatability of two successive binaural threshold determinations. Ten of the test subjects were participating in sound attenuation measurements for the first time, the other subjects had experience from earlier studies. 10 of them had taken part in tests more than 10 times. The test site fulfilled the requirements of /1/. The hearing threshold measurements were performed by means of computer-controlled audiometric equipment using the bracketing method according to ISO 6129 with a 2 dB step-size.

In each session first the threshold of audibility was determined with one of the two hearing protectors in place, followed by the threshold with uncoupled ears and the threshold with the second hearing protector being worn. The position of the protectors was varied from subject to subject. The initial fitting of the hearing protectors was carefully performed by the test subjects and checked by the experimenter. The adjustment was finally optimized by the subject in the presence of a broad-band noise. Each test session took about 45 minutes. The subsequent evaluation of the test results showed that the order of the hearing protectors tested during each session did not influence the result, i.e. on the average the hearing threshold levels of the test subjects neither improved due to learning effects nor were they impaired by fatigue.

In the more recent experiment B the sound attenuation provided by the two hearing protectors on test subjects has also been measured using a subminiature microphone. This microphone (Sennheiser KE 4-211-5) was placed in a stable position about 5 to 10 mm into the ear canal of the subjects. Its surface covered less than 50 % of the area of medium ear canals. The necessary wires were only 0.3 mm in diameter. A substantial influence on the performance of the hearing protectors can, therefore, be excluded. Test site and test signals were the same as before with the exception that the signal level was 75 dB for each frequency band. The tests were performed using 22 subjects from the whole group of 30 subjects used in experiment A, who were still available at the time of experiment B.

After the placement of the microphone, first the ear canal sound pressure levels were measured with one of the protectors in place followed by the equivalent measurement without the hearing protector being worn. This sequence was repeated twice, followed by the same procedure with the second protector. Each time the hearing protector was carefully adjusted by the test subject in the presence of a broad-band noise. Each test session took about 30 minutes.

3. RESULTS

3.1 ISO 4869 Versus Microphone Measurements

The mean sound attenuation and the corresponding standard deviations were calculated for each of the two protectors on the basis of 22 test subjects each being tested three times in each of the two experiments. They are presented in the figure for one of the protectors (Silenta Super).

There is considerable, good agreement between the results of the two experiments in the frequency range from 250 Hz to 4000 Hz with a maximum deviation of ±1.1 dB at 1000 Hz, confirming that sound attenuation data according to ISO 4869 may be regarded as also valid for the application of hearing protectors at higher sound pressure levels. Below 250 Hz, sound attenuation measurements according to ISO 4869 are apparently affected by the presence of physiological noise, the results from microphone measurements being therefore more realistic. This finding is in good qualitative agreement with experimental results of other authors /3/. It is worth noting that the mean sound attenuation according to ISO 4869 for the subgroup of experienced listeners was clearly lower at these frequencies than for the subgroup of inexpe-
rienced listeners. Even within the group of inexperienced listeners the measured sound attenuation was lower for the third measurement than for the first. This indicates that the ability to detect a test signal in the presence of masking noise can be increased by experience.

Sound attenuation measurements with a microphone technique cannot include the limiting effect of bone conduction which very often is an important factor at 8000 Hz and sometimes - depending on the attenuation of the protector under test - at lower frequencies down to 1000 Hz. At these frequencies the result according to ISO 4869 may be regarded as true.

The results are quite similar, in principle, for the second protector tested though the differences between the results of experiments A and B are a little higher in the frequency range 250 Hz to 4000 Hz (up to +1.4 dB at 250 Hz and -2.3 dB at 4000 Hz). Summarizing, it can be stated that ISO 4869 principally yields accurate sound attenuation data in the most important frequency range, while only the figures at the lowest frequencies may be overestimated, depending on the type of hearing protector and the experience of test subjects.

3.2 Repeatability

Individual test subjects showed average maximum deviations between 3 independent sound attenuation measurements according to ISO 4869 of about 5 dB for each test signal. Two sources usually contribute to this spread of data, i.e. the uncertainty of hearing threshold level determinations and the actual fitting of the hearing protector. The first factor was determined separately by uncorrelated hearing threshold level measurements in direct succession resulting in an average maximum deviation of about 3 dB between 3 independent measurements. The influence of the fitting could be extracted from the microphone measurements and was about 2 dB on the average for three independent trials.

Groups of 11 test subjects, on the other hand, showed a smaller spread of the test results for repeated measurements. However, even in this case deviations of the group means of up to 2.5 dB may occur at any frequency using the microphone technique, while 5 dB may occur in measurements according to ISO 4869. The group of experienced listeners did not produce essentially more reproducible results than the group of inexperienced listeners. However, this might be different in the case of hearing protectors where it is harder to achieve a proper fit, e.g. ear plugs.

Finally, the group of 22 test subjects showed deviations of up to 2 dB between repeated measurements in each of the two experiments.

3.3 Inter-Subject Variability

In order to study the influence of some evident test subject characteristics on the sound attenuation and to get a quantitative measure of the possible spread of attenuation data when the tests are performed with a different group, a number of subgroups of 6, 10 or 15 test subjects each have been extracted from the total group of 30 test subjects used in experiment A. Some of the selection criteria used were head width, head height, hair state, hearing threshold level, experience in sound attenuation tests. Though a detailed statistical analysis has not yet been performed a slight increase of sound attenuation with increasing head width, increasing head height and decreasing hair density around the pinna indicates that it might be wise to aspire to balanced groups of test subjects with respect to such criteria. Moreover, it transpired that the group of most sensitive listeners produced the highest attenuation data (compare /4/ for a possible explanation of this effect).

Subgroups of subjects were also selected on the basis of extreme sound attenuation. The resulting group means indicate the maximum differences that might occur at any frequency when hearing protectors are tested with different groups of subjects in one laboratory. The outcome of this evaluation was maximum deviations of 13.5 dB (mean of 6 subjects), 11 dB (mean of 10 subjects) and 8.5 dB (mean of 15 subjects) when the groups are tested once. These figures decrease by about 2.5 dB when the groups are tested three times each. The maximum deviations usually occurred at the test signal centred at 8000 Hz. Thus, the inter-subject variability is considerably larger than the repeatability. This result is also qualitatively confirmed by the outcome of experiment B. As can be seen in the figure, the average standard deviation at medium frequencies is almost the same in both experiments, though the repeatability of microphone measurements is much better for single subjects.

3.4 Overall Variability

A measure of the deviations of mean sound attenuation that may occur in one laboratory using ISO 4869 which is more sophisticated than maximum deviations in the standard deviation of means calculated for various groups of test subjects. This information could be easily derived from a statistical treatment of the data for the various subgroups if the result for the total group, i.e. 30 test subjects, 3 trials each, is taken as being free from random uncertainty. The outcome of this calculation is given in the Table below.

<table>
<thead>
<tr>
<th>Average</th>
<th>VT</th>
<th>AVT</th>
<th>VAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.5</td>
<td>1.2</td>
<td>1.7</td>
<td>1.1</td>
</tr>
<tr>
<td>11.5</td>
<td>1.5</td>
<td>1.8</td>
<td>1.3</td>
</tr>
<tr>
<td>12.5</td>
<td>1.8</td>
<td>2.2</td>
<td>1.6</td>
</tr>
<tr>
<td>13.5</td>
<td>2.1</td>
<td>2.6</td>
<td>1.9</td>
</tr>
<tr>
<td>14.5</td>
<td>2.4</td>
<td>2.9</td>
<td>2.2</td>
</tr>
<tr>
<td>15.5</td>
<td>2.7</td>
<td>3.3</td>
<td>2.5</td>
</tr>
</tbody>
</table>

These results clearly show that the measurement uncertainty may be more effectively reduced by increasing the number of test subjects than by repeating measurements with the same small group. Testing 15 subjects once takes almost half the time of testing 10 subjects 3 times, the measurement uncertainty being almost equal in both cases.

4. CONCLUSIONS

This study has confirmed that the standardized method for sound attenuation measurements yields accurate results in the whole frequency range except at very low frequencies. The variability of the test results is, however, large when using small groups of test subjects, even in the case of custom-molded hearing protectors. Since the inter-subject variability clearly dominated the repeatability, a minimum number of 15 to 20 test subjects should be prescribed in future revisions of the standard in order to increase the comparability of test results.

HEARING PROTECTOR TESTING - AN EEC INTER-LABORATORY COMPARISON
M.S. Shipton
National Physical Laboratory
Queens Road, Teddington, Middlesex, TW11 0LW, UK.

INTRODUCTION

Over the years, various standards for the measurement of the attenuation characteristics of hearing protectors have been adopted by different countries. In 1983 much of this work was brought together with the publication of an ISO standard (1). A similar test procedure had been used for some years in several different countries but no formal inter-comparison had been carried out. As the procedure uses a number of test subjects to obtain estimates of the attenuation, the test is time consuming and therefore expensive. Recognising this, the ISO working group responsible also devised an objective test (2). It was recognized that the insertion loss measured by this method does not directly relate to the attenuation of the protector but the method was thought to be suitable for quality-control purposes.

The objective of the work described in this paper was to compare the results of attenuation and insertion loss measured in different European test laboratories and, where possible, to resolve any significant differences. In order to save test time, a radial rather than a round-robin procedure was adopted. The muff types chosen were: Rilsom Blue, Optac Optigard 4000, Racial Safety Sonogard and Willson 328. In addition, each laboratory received sufficient samples of EAR foam-plastic earplugs to enable the subjective test to be carried out on this plug. These protectors were chosen to be representative of commonly available types and to encompass the major differences in construction, ex plastic and metal headbands, foam- and fluid-filled ear cushions.

The central reporting laboratory made measurements on numerous samples of these protectors by both the objective and the subjective test procedures and a selection of each type was sent to each participating laboratory for evaluation before being returned to the central laboratory for re-measurement.

The National Physical Laboratory (United Kingdom) acted as the central reporting laboratory, the other participating laboratories were as follows:

- Danmarks Tekniske Højskole, Denmark
- Institut National de Recherche et de Securite, France
- Physikalisch-Technische Bundesanstalt, Federal Republic of Germany
- TNO Research Institute for Environmental Hygiene, Netherlands

A detailed report of the findings is to be published and only a summary of the more significant results is given here.

TEST PROCEDURES

The procedure for the assessment of protector attenuation (subjective test) is to determine the free-field threshold of hearing of a number of test subjects both with and without the protector in position. The attenuation of the device is then taken to be the mean difference between the two threshold values, in decibels. A similar technique is used in the objective test method for the measurement of insertion loss, but in this case the test subject is replaced by an acoustic test fixture (ATF) which incorporates a microphone; the difference between the two measured sound pressure levels is taken to be the insertion loss of the muff. Note, however, that the objective test method is applicable only to ear muffs.

Prior to the start of the intercomparison all participants met to discuss measurement procedures to ensure their feasibility, for example, differences would reflect true inter-laboratory differences rather than mere procedural differences. For instance, it was agreed that, for subjective testing, each laboratory would use 21 test subjects rather than the minimum of 10 allowed by the standard.

The standards for both test methods closely define the sound field at the position occupied by the subject's head or the ATF but, in the case of the objective test field such field can be achieved in an acoustic duct, a random incidence field or in a plane progressive field.

OBJECTIVE TESTS

Each participating laboratory used its own implementation of the ATF to make objective measurements on four samples (both shells) of each of the four muff types. Measurements were made in rooms whose acoustical characteristics ranged from free field to reverberant but all sound fields complied with the standard.

All the muff types were tested by the central laboratory both before and after circulation to participants. For each muff type, small but systematic differences between these two measurements were noted, the frequency dependence being different for each muff type. Averaged over all test frequencies and muff types, the mean difference was 0.7 dB; only one muff type showed a mean difference across frequency in excess of 1 dB. Some of this variation can probably be attributed to ageing of the headbands, since muff types were handed-new when first tested by the central laboratory.

Repeat measurements made by an individual laboratory on the same muff sometimes showed widely differing results. Although each determination of insertion loss showed that the average of a number of repeat measurements, these repeat values could vary by several decibels. Direct comparison between laboratories is difficult since each laboratory tested a different set of muffs. Therefore, only a coarse comparison between laboratories has been possible, using the average values of results from all laboratories for a particular muff type as the datum line (see Figure 1); typically laboratories were within 3 dB of this datum line. However, one laboratory showed muffs which tended to lie at an extremity and this deviation was traced to faulty assembly of their ATF.

SUBJECTIVE TESTS

Each laboratory tested one sample of each of the four muff types and the one type of earplug using the subjective method. For each muff type, the test muffs were selected by using the objective test results from the central laboratory. The selected muffs were then tested subjectively by the central laboratory before being sent to the other laboratories. Subjective results from the central laboratory showed that the muffs were not significantly different, the between-muffs variance
of the muff attenuations at any one test frequency being typically 2 dB. Comparing results from all laboratories, this same between-muffs variance increased to 13 dB and the attenuation obtained by the participating laboratories and the central laboratory differed by as much as 6 dB. Comparing data averaged over all muff types, some laboratories showed persistent trends above or below the average over all laboratories (see Figure 2). At 2 kHz and 4 kHz where the discrepancy was large, this has been traced to a laboratory oversight whilst at 63 Hz the discrepancy is thought to be caused by loudspeaker distortion.

It is assumed that all laboratories had similar plugs, since they were all taken from the same batch. Prior to testing, laboratories agreed a protocol for fitting the earplugs but, in spite of this, differences in measured attenuation between laboratories of up to 10 dB were observed, with a typical maximum difference averaged over all frequencies of about 9 dB. Fit appears to be the most significant factor, since repeated measurements by the central laboratory showed that shifts of 2 dB were possible by even slight variation in the fitting procedure; using the same group of test subjects and the same fitting procedure, however, a high measure of repeatability was achieved.

CONCLUSIONS

Results obtained on the ATF are variable and it is essential that measurements should not be based on a single reading. Direct comparison of data from the objective tests was made difficult by the fact that (a) no muff was common to all laboratories and (b) that different test fields are allowed by the standard.

Using the subjective test method it is possible for any one laboratory to obtain good repeatability but substantial differences between different test laboratories can be expected.

REFERENCES


NORDIC ROUND ROBIN TEST ON HEARING PROTECTOR MEASUREMENTS

Turben Poulsen

The Acoustics Laboratory, Building 352, Technical University of Denmark, DK-2800 Lyngby, Denmark

INTRODUCTION

A few years ago the Nordic countries agreed upon a common standard for hearing protector requirements. The aim of introducing a common standard was to make a certification of a hearing protector in one country valid in the other countries too. According to the attenuation values obtained by a standardized measuring method a hearing protector could be rated into one of three classes. Furthermore, the standard specifies some minimum requirements to heat and cold resistance, and stability of a headband's elasticity, fire resistance, corrosion, etc. Only acoustical results are included in this paper.

The acoustical test methods are essentially the same as the ISO 4869 standard and the Draft ISO Standard ISO/DIS 6290. As these test methods were rather new it was felt necessary to perform an intercomparison among the Nordic Testing Laboratories before the certification procedure was put into force.

Round Robin Test

Four testing laboratories participated in the intercomparison:
- Dept. of Technical Audiology, Stockholm, Sweden
- Audiological Laboratory, Oslo, Norway
- Dept. of Industrial Hygiene and Toxicology, Vaasa, Finland
- The Acoustics Laboratory, Lyngby, Denmark

The same set of hearing protectors was measured in turn by these four laboratories, and this procedure is called a Round Robin test. The hearing protectors were measured both before and after circulation by the coordinating laboratory in Denmark.

Four earmuff types (two specimens of each) and two earplug types were selected for the test:
- Sonegard. Liquid-filled cushions. Steel headband.
- Silenta Super. Foam-filled cushions. Steel headband.
- Propp-u-plast. Glass wool covered by a thin layer of plastic.

SUBJECTIVE MEASUREMENTS

The attenuation of the hearing protectors is determined by a hearing threshold procedure. Two thresholds are measured, i.e. with and without the protector in place, and the attenuation is determined as the difference between the two thresholds. The thresholds are determined for 1/3 octave noise bands with centre frequencies from 125 Hz to 8000 Hz. The sound field is approximately diffuse at the position of the test subject's head. 20 test subjects participated at each laboratory.

The sequence of frequencies and sequence of thresholds were the same at all laboratories in order to reduce the random variability between laboratories.

Results. Subjective Method

The common Nordic standard contains requirements to acoustical attenuation based on the lower quartile of data from 20 test subjects. Comparison of results on the basis of quartiles raises certain theoretical problems for the statistical analysis. Therefore the mean values have been used as a basis for the comparison.

Fig. 1 shows a typical example of the spread between the muff results from the four laboratories. Both mean values over 20 test subjects and the corresponding standard deviations are shown.

Fig. 1 Typical subjective earmuff results. Mean and standard deviations from 20 test subjects at 4 different laboratories.

The maximum differences are generally around 4 dB, and the standard deviations are typically around 3-4 dB with a maximum around 6 dB.

Fig. 2 shows the results from one of the plug measurements.

Fig. 2 Subjective earplug results. Mean and standard deviations from 20 test subjects at 4 different laboratories.

Again, the mean-value curves show the same general course with differences up to about 6 dB. The standard deviations are typically around 7 dB with a maximum around 9 dB.

In order to facilitate the comparisons typical 95% confidence intervals are shown in Figs. 1 and 2.

OBJECTIVE MEASUREMENTS

The objective method uses an acoustic test fixture instead of test subjects. The test fixture is
equipped with a microphone and two end faces resembling the cup separation obtained with an average human head. With pink noise the 1/3 octave spectrum from 63 Hz to 8000 Hz is measured in a sound field with and without the protector placed on the fixture. The difference between the two spectra gives the insertion loss for the hearing protector. This method can be used for muff-type protectors only.

Results. Objective Method

A typical result is shown in Fig. 3. The curves shown are mean values over the 4 cups as measured at the four laboratories. Standard deviations are shown too.

![Graph showing mean insertion loss](image)

**Fig. 3 Typical objective earmuff result. Mean and standard deviations over 4 cups measured at four laboratories.**

The differences between laboratory results are generally within 3 dB. The standard deviation increases with decreasing frequency. This is seen for the two protectors with liquid-filled cushions, but is not found for the two protectors with foam-filled cushions.

**DISCUSSION**

Comparison of results obtained at the coordinating laboratory before and after the circulation showed that the protectors have not changed their performance during the tests.

Subjective Results

A statistical analysis of the subjective results showed significantly different attenuation values between laboratorties for the muff-type protectors, but not for the plug-type protectors. This means that the four curves shown in Fig. 1 are statistically different, whereas the curves in Fig. 2 are not, although the separation between curves are greatest in the latter figure. This is due to the fact that the variance of the individual data is much greater for the plugs than for the mufffs, and thus also the confidence limits shown in the figures. The plug variance may be reduced by a better fitting procedure.

The statistical analysis showed, furthermore, significant interactions in the material, preventing application of a laboratory correction. The residual variance in the subjective material showed that the measuring error is less than about 3.5 dB for the mufffs and less than about 4.5 dB for the plugs.

Although significant, the muff attenuation differences between laboratories are small and only important if the results are compared uncritically with very strict classification limits. The variance associated with this laboratory effect is 6-10 times less than the variance originating from the 20 test subjects. It may thus be more efficient to measure with a greater number of test subjects at a single laboratory than to measure at many laboratories.

**Objective Results**

A statistical analysis of the objective data showed significantly different insertion loss values between laboratories. The insertion losses - given as mean values over 4 cups - differ in some cases by more than 6 dB. Such great variations were not expected, as the method is a purely objective one without any test subjects involved.

An analysis of the results from the individual cups showed that both main effects and interactions are all significant. This means that a laboratory correction cannot be applied to the objective data either. From the residual variance the measuring error may be estimated to about 2 dB.

Due to the great variances the objective method cannot be used as a reproducible control method. The reason for the bad reproducibility of this method is not known at present, but a systematic investigation of this topic will take place in the Nordic countries in 1986.

**CONCLUSION**

The Round Robin Test has shown that
- the inter-subject variance for the earplugs is not significantly different.
- the inter-subject variance is considerably less for mufffs than for plugs, and thus a laboratory effect is found for mufffs.
- the laboratory variance is 6-10 times less than the test subject variance.
- concerning class limits the data shall be used with suitable regard to the random measuring error.
- the random measuring error is about 4 dB for the subjective measurements.
- the fitting procedure may be critical for some plugs.
- unexpected big differences were found in the subjective insertion loss results.
- the objective method cannot - in its present form - be used as a control method.
- a systematic investigation on the origins of the objective variability is needed.

**REFERENCES**


MEASUREMENT OF LEVEL-DEPENDENT ATTENUATION IN HEARING PROTECTORS

C. H. Allen* and E. H. Berger†

*The Clayton H. Allen Corporation, 80 South Road, Chebeague Island, Maine 04017-9710
†E-A-K Division, Cabot Corporation, 791 Zionsville Road, Indianapolis, Indiana 46268-1650

INTRODUCTION

Conventional hearing protection devices (HPDs) provide an amount of attenuation that is independent of the incident sound level. This is beneficial in general, but may be less than optimal in special situations where higher or lower levels of protection are desired.

By contrast, a level-dependent HPD is defined here as a conventional hearing protector that is modified to bypass the ear when the sound level is low, and to attenuate the sound when the level is high. This provides a more adjustable level of protection.

A level-dependent HPD has three ranges of performance:

1. The low-level range, in which the HPD provides minimal attenuation; this is typically the condition for which the HPD is designed.
2. The high-level range, in which the HPD provides the maximum attenuation; this is typically the condition for which the HPD is designed.
3. The intermediate range, in which the HPD provides a moderate level of attenuation.

The practical advantage of a level-dependent HPD is that, during intervals of low noise level, it offers not only improved comfort but also reduces the overall impact of noise on the user.

ATTENUATION MEASUREMENT

We have demonstrated that conventional HPDs, indeed, are level-independent. Measurements using an acoustical test fixture (ATF), including a simulated flesh covering, were validated at sound levels below 90 dB, by comparison to the reference standard ear-attenuation at threshold (REAT) paradigm. The sound levels used were from 90 - 155 dB (broad-band steady state) using acoustical headphones and/or earplugs, and from 0 to 160 dB peak levels, using impulsive noise. Insertion loss (IL) was found to remain constant within 2 - 4 dB over that range of levels.

Since the IL of a level-dependent HPD is level-dependent, the required sound levels and attenuation increases beyond a predetermined level. The bypass is automatically reduced so that the attenuation increases with level and approaches that of the unmodified HPD.

Above the transition level, the IL of a passive bypass orifice increases at a rate approaching half the rate of increase in the incident sound level. The IL also varies with frequency, and can be affected by the acoustic characteristics of the path encountered inside the ear muff. Three methods of measuring the IL in the intermediate range have been used:

1. For evaluating the performance of a level-dependent HPD, the element alone, sound levels covering the entire range of interest up to 170 dB or higher can be produced by a pressure source in a small cavity coupled to the test item.
2. For evaluating the performance of a complete earmuff assembly, the direction of incidence becomes an important variable, so a simple pressure source is not suitable. Either a duct, a reverberation chamber or a semi-anechoic room are needed where a specifiable sound field can be generated. Generally steady-state level, or an appropriate level, is used for such tests, but the IL of the HPD can be measured at one or more levels by positioning the ATF inside an acoustic fatigue test chamber driven by an air-modulator capable of delivering the order of 10,000 acoustic watts.
3. For evaluating impulse response, a 22 calibre starter pistol can be used to generate progressive shock waves with peak levels, controlled between 130 and 160 dB or higher by varying the distance between the source and the ATF.
MEASURED RESULTS

The pure tone performance, that is typical, of an isolated passive level-dependent bypass is shown in Fig. 1, for one frequency. The IL is seen to be constant at low incident levels, but begins to increase near the transition level (here, 115 dB) and approaches an asymptote of 5 dB IL / 10 dB SPL.

When a bypass is combined with an earplug, similar performance is observed, but the increase in IL is limited to about one-third that of the unmodified earplug as illustrated in Fig. 2, where the entire transition range was observed using a small pressure source and an appropriate ATF.

When a bypass is combined with an earmuff into a reverberation room as the sound level may typically be near 20 dB to furnish an appropriate minimum of protection against background noise. It then provides increased protection against impulsive and other irregular peak noise intervals for noise above noise. (Fig. 3)

Representative IL spectra, as measured by low-level procedures with the bypass open and closed, are presented in Fig. 3, along with a third spectrum for the unmodified earmuff. This clearly shows the range over which the bypass controls the IL. When the bypass is closed the IL does not exactly equal that of the unmodified earmuff. The residual difference in the range 500 to 3000 Hz has been related to solid-borne vibration transmitted through the duct and earpad; the increased IL above 3000 Hz is due to shielding of the ear by the earpad.

The transition level and the initial rate of change in IL have been observed with an ATF in a reverberation room as the broad-band continuous noise level was raised to its limit of 120 dB. Figure 4 shows a typical result in the 500 Hz, 1/3-octave band. It resembles the similar performance of an isolated bypass element, Fig. 1. The transition level is seen to be well determined at an incident level of 117 dB rms. Figure 4 also shows the same frequency band, measured with impulses from a blank cartridge fire at various distances. The asymptotic low-level IL for both continuous and impulsive signals agree. The impulse data are plotted against peak levels so the transition level appears to be higher, but replottting the impulse data against rms levels (using an effective peak-to-rms ratio of 10) causes a lateral shift of 20 dB which brings the two sets of data into agreement on transition level also.

The range of level for the impulse measurements was great enough (160 dB peak level; 140 dB rms level) to show that the rate of increase in IL came close to the expected asymptote, then crossed it and leveled off at an upper limit (near 30 dB) which is found to be set by the IL of the overall structure of this earmuff assembly at 500 Hz.

The IL for broadband noise (.5 to 4 kHz) was followed from low levels through the transition level up to 145 dB, using an ATF positioned in the horn of a driver. The results were consistent with those from the methods just discussed and were used to determine the intermediate spectra shown by dashed lines in Fig. 3. The intermediate spectra determined by such high-level measurements on a complete earmuff have been found to agree with those derived from the low-level IL spectra of the earmuff, with the bypass open and closed, by combining them with high-level spectra of its isolated bypass element.

The horn system is a convenient tool for studying details of the level-dependent IL spectra. For example, when a level-dependent HPD was exposed to broad-band noise, the change in IL for an individual 1/3-octave band of that noise was found to be related to the overall incident sound level, not to the level in the narrow band itself.

CONCLUSIONS

Whereas a conventional HPD has a single IL spectrum, which is independent of the incident sound level, a level-dependent HPD has two. They characterize its operation at low and at very high ranges of incident sound level, respectively, and can be evaluated by conventional methods, at low sound levels, by measuring the HPD, first operating normally, and again, with the level-dependent assembly present, but blocked. The IL in the intermediate range can be found by suitably combining these IL values with those for the isolated level-dependent element. Preliminary results, obtained from tests in high-level sound fields, support the validity of this procedure and indicate that the measurement of IL for complete level-dependent HPDs at incident sound levels above 130 dB, generally may be unnecessary.
REAL-WORLD ATTENUATION OF HEARING PROTECTORS

G.F. Snoorenburg, R.H. ten Baa and A.M. Mimen
TNO Institute for Perception, P.O. Box 23, 3769 DE Soesterberg, The Netherlands

INTRODUCTION

Nowadays the measurement of hearing protector attenuation is well standardized. Attenuation figures determined according to the ASA standard 22.4.22-1957 are available for virtually all hearing protectors. The standard procedure is based on subjective measurements of the hearing threshold with and without hearing protector. The hearing thresholds are determined with a loudspeaker in front of the observer in an anechoic room using sinusoidal stimuli. Since strictly frontal incidence of the sound waves and sinusoidal stimuli are not representative of most practical situations, a new standard (EAS 4860-1981) is based on (quasi) random incidence in reverberating rooms using narrow-band noise stimuli. Detailed instructions about measurement procedure and measurement accuracy accompany the attenuation figures. Compatible data obtained in different institutes and, with the new IISO procedure, attenuation figures that apply to most practical situations.

The attenuation figures, however, are obtained in laboratory situations in which much attention is paid to proper insertion of earplugs and proper placement of earmuffs. In real-world labour conditions much less attention may be paid to proper fitting of the hearing protector and a proportion may require more effort. In particular for earplugs, disappointing real-world attenuation figures have been reported. For the V51R earplug Padilla (1978) found for 443 people in the real-world situation an average attenuation of 17-18 dB with a standard deviation of 8-9 dB whereas the laboratory figures were 25 dB and 5 dB, respectively. Edwards et al. (1978) found for 168 people working in six different factories that the attenuation of the single-flange V51R (a triple-flange earplug and a plug consisting of Swedish wool) was only 32 to 51% of the dB-value found in the laboratory; a difference of 13 to 18 dB.

Below, we shall describe our real-world results for the Wilson EP100 twin-flange earplug. The data were collected in 1979 but they were not fully reported. Summaries of the data were presented before in Snoorenburg (1982a, 1982b).

REAL-WORLD ATTENUATION OF A TWIN-FRANGE EARPLUG

The investigation was carried out among 98 servicemen (96 ears) of the Dutch Air Force. The Wilson EP100 earplug in part of the standard outfit of servicemen in the Dutch Forces. At the beginning of the training programme the servicemen were instructed on how to insert the earplug. No instruction was given at the day of the investigation. This day was within three months from the day they received the instruction. The subjects were not aware of the planned measurements. Before they were to start shooting they were ordered to place an earmuff over their ears for additional protection. This prevented the subjects from touching the earplug until the time of the attenuation measurement. The subjects were recruited before shooting commenced.

The subjective hearing threshold was measured with an automatic up-and-down audiometer (Interacoustics BA 2). To exclude all possible interference from environmental noise and to keep the audiometric telephones free from the effect of the telephones (Beyer DT485) were mounted in a big ear muff (Noiseol Mark II). The earmuff placed at the shooting range was first carefully removed making sure that the earplugs would not be moved. Next, the earmuff provided with the telephones was placed over the ear, again making sure that the earplugs would not be moved. Thereupon, the hearing thresholds for the right, the left and again the right ear with the plugs inserted were measured at 0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz. This was followed by measurements of the hearing thresholds of successively the left and right ear without earplugs.

The right ear provided with the plug was measured twice in order to get some insight into a possible learning effect in the audiometric procedure. Since the ear-with-plug condition always has to be measured first a learning effect in the audiometric threshold (lower thresholds as the procedure progresses) may give, after subtraction of the ear-without-plug thresholds from the ear-with-plug thresholds, attenuation values that are too high. The thresholds found in the repeated measurement appeared to be systematically lower than those in the first measurement. The effect, 0.7 to 2.9 dB depending on frequency, is small, however, and since experience has taught us that most of the learning effect resides in the first threshold determinations we may expect unbiased attenuation values when the first series is taken as a practising series only, not using the data. Measurement of earplug attenuation with telephones mounted in an earmuff does not produce results that are quite comparable to the standardized free-field methods described in the introduction. Further analysis has shown, however, that our closed-field procedure did not effect our results in any essential way.

The results for 500 Hz (typical of the results for 250-1000 Hz) and for 3000 Hz (at which the highest attenuation was found) are presented in Fig. 1. The results concern only 92 ears, three earplugs dropped out of the ear (unnoticed) and one was inserted inside-out. Fig. 1 shows median attenuation values of only 3 dB at 500 Hz and 23 dB at 3000 Hz. The mean values were about 4.5 dB.

![Figure 1](image-url)
higher. The variability in attenuation is considerable. At 500 Hz there was no attenuation at all in 25% of the ears while attenuation values in excess of 33 dB were found in 5% of the cases. At 3000 Hz there was only 3 dB or less attenuation in 5% of the ears and 54 dB or more in another 5%. At the successive frequencies from 250 Hz to 8000 Hz the average attenuation was 6.8, 7.7, 9.0, 19.4, 27.3, 24.2, 19.2 and 15.2 dB while the standard deviation was 9.7, 12.5, 12.9, 14.9, 14.9, 14.1, 16.2 and 18.6 dB, respectively. The smooth curves in Fig. 1 show that the distribution of attenuation values can be described in first approximation by a normal distribution of the dB-values. The normal distribution is determined by the average dB-value and the standard deviation.

Fig. 2 shows the attenuation exceeded in 5, 10, 25, 50, 75, 90 and 95% of the cases, respectively. The negative attenuation values at the higher percentiles must be ascribed to measurement inaccuracy. For comparison, we included some data into this figure that were collected by us on several occasions (3 subjects) in the laboratory according to standard 224.22. In contrast to the previous data, these data are based on hearing threshold measured for two ears simultaneously. Fig. 2 shows that the mean laboratory attenuation lowered by one standard deviation is exceeded by only about 25% of the real-world cases. The mean laboratory value minus two standard deviations roughly matches the median real-world value. Thus, the laboratory attenuation values are deceptively high. We should realize, however, that the real-world figures from the present study may be on the low side. With shooting noise there is no good opportunity to check the seal when the earplug is inserted whereas continuous noise provides this opportunity.

**Fig. 2.** Real-world attenuation of the EP100 as a function of frequency exceeded by the indicated percentage of 92 ears. The two shaded curves represent the average laboratory value minus one and minus two standard deviations, respectively.

**ASSUMED ATTENUATION VALUES**

The variability in hearing protector attenuation is too great to be ignored. Therefore, so-called assumed values are proposed for practical purposes which take this variability into account. It is interesting to note that assumed values, equal to e.g. the average value minus one or two standard deviations, are being proposed irrespective of accepted damage risk criteria. When it is accepted that a certain percentage of the population exposed should not develop a certain hearing loss or more, this should also hold for people using hearing protection. The effective attenuation value can then be derived from this criterion without making any further choices. When, on the other hand, the variability in attenuation is mainly due to variability in placement of the hearing protector within the ear canal, the equivalent noise exposure should be calculated taking this variability into account. The effective attenuation becomes the difference between this protected noise exposure level and the unprotected exposure level. When, on the other hand, the variability is mainly due to intersubject differences in the fit of a hearing protector the resulting intersubject variability in noise exposure should be combined with the intersubject variability in susceptibility to noise.

A pilot experiment in the laboratory carried out according to ISO 4869 showed for the EP100 earplug that the intersubject variability is much greater than the intrasubject variability in earplug placement. (Only one ear was measured, the other ear was fitted with an earplug and an earmuff.) Below 1000 Hz 70% of the variance is intersubject variability. 25% is intrasubject variability and 5% is measurement error. Above 1000 Hz these values are 85, 10 and 5%, respectively. Thus, the intersubject variability is most important and the variability in attenuation should be evaluated primarily with respect to variability in the exposure level at which a subject develops a certain hearing loss. This variability in exposure level depends on the type of noise; it is greater for impulse noise than for continuous noise. A likely value would be a standard deviation of 6 dB. When the standard deviation in attenuation also equals 6 dB, the total standard deviation in the protected exposure level at which a subject develops a certain hearing loss increases to 6/2 = 8.5 dB. When the damage risk criterion is based upon protection of the population at e.g. the 10% level (1.28 standard deviation from the mean value assuming normal distributions) this implies an extra offset in the exposure level of -1.28 x (8.5 - 6) = -3.2 dB in order to reach the 10% level. Thus, in this example the effective attenuation is 3.2 dB lower than the average value.

**REFERENCES**


SELECTION OF HEARING PROTECTORS

A. Behar and R.A. Jackson
757 McKay Road, Pickering, Ontario, Canada L1W 3C8

INTRODUCTION

When selecting a hearing protector, the user decides first on the type he will be using: cup-mounted, plug or nasal plugs. At this time, his only concern is the type of work being
performed and the work environment.

Next, he usually looks upon the Noise Reduction Rating (NRR) assuming that the higher
the NRR, the better the protector. In doing so, he does not take into account the fact that the
NRR reported by the manufacturer is the highest ever to be obtained under ideal condition.
The comfort that the user will experience is never considered since it is not included in
the manufacturer's literature. However, if the old saying by Glorij, “The best protector
is the one that is worn” is still valid, then every effort should be made so that comfort be
included in the process of selection.

Since our Corporation was to update the List of Approved Hearing Protectors, we were
requested to provide criteria for selection that will take into account both characteristics:
acoustical attenuation and comfort. This paper describes the method we developed for that
purpose. Although it can be applied by any user of hearing protectors, its main use is for large industries where workers are exposed to a variety of noises of different durations, sequences and/or spectra.

ACOUSTICAL ATTENUATION

Acoustical attenuation is measured in specialized laboratories following the procedures
in the ANSI Standard S3.19-1974. Results are presented as tables of mean attenuation and
standard deviations at nine different frequencies. The NRR is calculated from these results. To
predict the noise level of the protected ear (under the protector), the attenuation is
subtracted from the octave band spectrum of the existing noise. Another way of doing the
same calculation is by subtracting the NRR of the protector from the noise level measured in
dBA or dBA.

NRR is often preferred instead of the
acoustical attenuation because it avoids measuring the noise level in octave band and simplifies
the calculation process. Finally, if the
noise exposure of workers is used for the
calculation instead of noise levels, only the
NRR can be used since dosimeters are set to
measure in dBA.

It has been reported repeatedly in the
literature that the real life NRR is much
smaller than its nominal value. Some researchers suggest that the difference is in the order of
10 dBA while others (including (NIOSH)) recommend the nominal value be divided by a factor of two.

In our case, we decided to do our own
testing with the intent of having more realistic results. Twenty-two protectors were measured
altogether in our facilities. The NRR was
subsequently calculated. However, we preferred to use the NRR 84, that is, the NRR for 84% of
the population. (NRR 84 is calculated by subtracting one standard deviation from the
mean attenuation value at each of the testing
frequencies.

COMFORT

No standards for testing comfort exist to our knowledge. As with other subjective
characteristics, comfort can be assessed by using questionnaires, therefore, we had to
design one for that purpose. It is used with the test where ten persons wear the protector
for as long as a full shift. They have to
answer three questions as soon as they have
the protector on and another three at the end of the test. Each question has five possible answers that have numerical scores. The result of the questionnaires is a single number obtained by adding the individual scores.
The questionnaires were tested with one of the protectors that has been in use in the Corporation for a long time.

CRITERIA FOR SELECTION

Attenuation

Noise exposure surveys are performed in our Corporation on almost 10,000 workers by using
sampling methods on trades (groups of workers performing the same tasks or working in the
same noisy area). Mean noise exposure levels
(\textit{L}^{\text{ eq}})\text{ at the upper 95\% confidence level,}
standard deviations and the number of workers in each group were determined. By dividing
the standard deviation to each \textit{L}^{\text{ eq}}, we obtained the
maximum noise exposure level that had 85% of
the workers from each group.
The next step was to calculate the number of workers with noise exposures exceeding a
given level. For instance, we found that noise exposures of some 7,000 workers exceed 90 dBC.
(One the spectral characteristic of most
noises in our plants, a 5 dB correction is
sufficient to convert dBA in dBC.) Consequently, a protector with NRR of 5 will reduce the noise
exposure of all but the above 7,000 workers to
an acceptable level of 85 dBA.

By using the same process, we found that a
protector with NRR 84 of 17 will bring noise exposure levels of 95% of all workers down to
85 dBA.

Consequently, we decided that the minimum
NRR 84 for our hearing protectors should be 17.

Comfort

Each answer in the comfort questionnaire was
coded from 1 to 5, 5 being the most favourable.
The highest possible score is 5. (If everyone chooses the answer \#5). The answer \#3 (neutral)
yields 3. For reasons of practicality, we chose
the middle point between answers \#3 and \#4, that is, a score of 3.5.

CONCLUSION

The method presented here takes into account both elements of a hearing protector: attenuation and comfort. It can be tailored to a noise exposure pattern particularly to a given industry that has to decide the level of protection needed (how many workers should have noise exposures below a certain level).
ATTENUATION MEASUREMENTS OF HEARING PROTECTOR IN WORKPLACE

A.M. Méndez, E.B. Salazar and H.G. Bonatti

Laboratorio de Acústica y Luminotecnia C.I.C.
Camino Centenario 506 - 1887 Goner - Argentina

INTRODUCTION

In industrial situations, when noise reduction is impossible, insufficient, or undesirable, it is advisable to use hearing protectors. One of the basic requirements to consider in the election of the protector, besides comfort, is attenuation. This attenuation is measured in dB and, in general, it is expressed in terms of frequency. There are many different methods to value the attenuation of a hearing protector, but not yet the experts in the topic have one definitive criterion in the selection of the best method of measurement. In the Argentine we have adopted the national standard IRAM 4060 to perform this measurement. This standard establishes the use of white noise filtered in third octave bands as test signal. The measure is performed in a reverberant room, to obtain diffuse field. The real attenuation response of the protector is, for one subject, the differences between the hearing thresholds, at each of the test frequencies, obtained with and without the protector. The measurements must be performed, at least to 10 subjects in 3 different times or to 15 subjects in 2 different times, completing 30 determinations. Before the test, the subjects must be instructed on the fitting of the protectors. In general, the standards establish measuring methods to determine the attenuation in laboratory conditions using new protectors, perfectly fitted up and with subjects that are not necessarily industrial workers. The object of this research was to evaluate the noise attenuation of different types of plugs and muffs, but considering also that we find in workplaces: the impairment of the elements by wearing; the improper fitting; the aging of the materials; etc.

INSTRUMENTATION

The measuring system was arranged by:

a) a white noise generator with a constant output (1 dB) in the range from 100 to 10000 Hz.

b) a third octave filter, type B&K 1616. Only the bands centered in the frequencies 125, 250, 500, 1000, 2000, 4000 and 8000 Hz were used.

c) one attenuator with a total range of 50 dB in steps of 1.5 dB. This range was extended by means of a switch that increased the signal in 30 dB.

b) a switch for the presentation of the signal. It was built up in base of one LDR resistor activated by a LED diode, so that it was possible to eliminate switching noises. The signal presentation is in the pulsed form with an on-off time of 1 second, in order to improve the definition in the hearing thresholds.

e) a 5 watt power amplifier built with an integrated circuit.

f) a loudspeaker, 15 cm in diameter, extended range type, mounted in its baffle.

The frequency response of the whole system is not critical, as the results are obtained by difference between the thresholds with and without protection, in each frequency band.

Fig.1: Instrumentation:
NG: white noise generator; F: B&K filter, type 1616; Att: 1.5 dB steps attenuator; Int: signal presentation switch; Amp: 5W amplifier; extended range loudspeaker.

MEASURING SITES

The tests were performed in one chemical and one automotive industry. In both plants we used the audiometric booths of the Health and Safety Division, adapted for these tests. The cabins were type IAC with very good insulation, set up in quiet places. The background noise inside each booth was measured with a Bruel & Kjaer sound level meter, type 2209, and one octave band filter type 1613. Both booths were perfect to perform audiometric tests, but their background noise exceeded, in some frequencies, the requirements of the standards. It is necessary to assume that perhaps it has been some masking effect in those frequencies. The results of the background noise measurements are shown in Table 1.

Table 1: Values of the background noise levels.

<table>
<thead>
<tr>
<th>Octave band</th>
<th>Plant</th>
<th>Limit #</th>
</tr>
</thead>
<tbody>
<tr>
<td>centre frequency</td>
<td>chemical</td>
<td>automotive</td>
</tr>
<tr>
<td>Hz</td>
<td>dB</td>
<td>dB</td>
</tr>
<tr>
<td>125</td>
<td>32</td>
<td>31</td>
</tr>
<tr>
<td>250</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>500</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>2000</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>4000</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>8000</td>
<td>13</td>
<td>12</td>
</tr>
</tbody>
</table>

# according to ANSI S3.19-1974

HEARING PROTECTORS TESTED

In the plants we made the research there are many different types of protector in use. For this reason, at first we measured nine different protectors but then only five were retained. We considered only those protectors, measured at least in 30 subjects, in order to obtain more reliable results, from the point of view of the statistical study. We measured four earplugs and one earmuff. All of them, with the exception of the Bilsom Propp-O-Plast plug, were made in Argentine protectors. There were two silicone rubber plugs, one premoulded plug and one muff. Excepting the Propp-O-Plast, that is a disposable plug, all the other protectors measured were those that in that moment were worn by the workers.

SUBJECTS

The subjects for this study were males, between 20 and 50 years old. All them had a recent audiometric recording that insured normal hearing threshold levels. The tests were performed early in the morning to avoid exposition to noise.
MEASURING PROCEDURE

The workers were asked to come to the test site with their own protectors. We started determining the hearing thresholds with protection and without it. (Fig.2) The subjects were not instructed in the protector fitting. Finally, the workers answered a questionnaire on different aspects about their protection elements.

Fig.2: Subject during the test.

RESULTS

In Figs. 3 to 7 graphs of attenuation and standard deviation are shown. We compare our results (continuous lines) with manufacturer's data (dashed lines). In general exists a clear difference in the attenuation results; they are lower in the field measurements. This fact has been reported by other authors, specially A. Behar, who has done a study similar to this (1). In standard deviation, the differences between values are not so important. As a conclusion it is necessary to suggest the importance of a good instruction to the workers in the use and care of protectors. It is necessary to improve the control of the plugs and muffs in order to replace the elements worn out by use.

ACKNOWLEDGEMENT

The authors thanks Mr. A. Behar (Ontario Hydro) for the advices and discussions during the fulfillment of this research.

Fig.3 Bilsom Propp-O-Plast Plugs

Fig.4 Saylens Plugs

Fig.5 Otosan Plugs

Fig.6 Premoulded Plugs

Fig.7 Norweg Muffs

REFERENCE

THE ATTENUATION OF INSERT EAR PROTECTORS IN MALES AND FEMALES

S. M. Abel, H. Kunov, P. W. Alberti and D. Nolas
Silversman Hearing Research Laboratory, Mount Sinai Hospital, 600 University Avenue, Room 843, Toronto, Canada MSG 1X5

INTRODUCTION

Industry-based hearing conservation programmes rely on personal hearing protection as a method for reduction of noise exposure that is efficient, reliable and cost-effective. Schemes for reduction of noise at the source may proceed in tandem but these are often difficult and expensive to implement, and may not always be feasible.

An important drawback of hearing protector usage is that the attenuation actually achieved in the workplace characteristically falls short of that specified by the manufacturer. In our own study, 350 workers were asked to bring to the clinic the protectors they normally used in the workplace and to fit them without further instruction. The results indicated that there was a wide variation both within and between a broad class of hearing protector types. It was not uncommon to observe a range in attenuation of 0 to 45 dB across individuals for a given protector and test frequency. The average attenuation was in many instances as much as 10 dB below the manufacturer's specification. This discrepancy could be reduced to 7 or 8 dB if the protector was fit by an audiologist.

Low achieved attenuation has been shown to be related to poor placement technique, improper fit, maintenance, deliberate abuse such as removal of flanges and poor hygiene. An important reason for misuse and deliberate abuse of protectors appears to be the decision to achieve greater comfort. According to Tengling and Lundin, 9 workers who wear protectors with relatively high attenuation generally have more severe hearing impairment than those who use protectors with 10-25 dB less maximal attenuation. In both groups, however, the preference for comfort is an anomaly that the better, less uncomfortable protector will be worn for a shorter duration making it relatively less effective over the longer term.

In a survey of protector utilization for 20 industrial sites in North Carolina, Rynster and Holder found that 32% of employees interviewed cited general discomfort and sore ear canals, 25% complained of ear canal irritation, infections and headaches, 14% mentioned improper fitting, and 35% noted structural failure of the device and difficulty in upkeep. The remaining 6% complained of interference of the device with communication. Our own research, as well as that of Rynster and Holder, suggests that females generally have smaller ear canals than males and thus, may have relatively greater difficulty in achieving a good seal and appropriate attenuation with those sizes of insert protectors currently available. The present study was designed to explore this possibility.

METHOD

A total of 120 subjects, 60 males and 60 females, 18 to 38 years of age and screened for a history of ear disease and hearing loss, were tested. Individuals were randomly assigned to one of three hearing protector groups, the only restriction being that each group comprise 20 males and 20 females.

The three types of insert hearing protector selected for study were the E-A-R expandable foam plug, the Wilson Sound Silencer premolded vinyl ear plug with double flange, and the Bilsom Soft polyethylene encapsulated glass fiber plug. All have similarly high NRR values. The E-A-R is compressed prior to insertion, and subsequently expands to fill the ear canal. The Wilson plug is pre-sized. The formable Bilsom plug like the E-A-R is available in one size only. In all cases, the protectors were inserted by a trained assistant.

For each subject the detection of one-third octave signals centred at 250, 500, 1000, 3150 and 6300 Hz was measured. The test signals, presented binaurally over headphones, had a duration of peak amplitude of 200 msec with linear rise and decay times of 100 msec. The ear canal was lined with a quiet background with the open ear and with ear plugs fitted binaurally. Subjects were tested individually in an IAC booth. The apparatus and two-interval, forced choice method used have been described previously. The ear canal was lined with a quiet background with the open ear and with ear plugs fitted binaurally. Subjects were tested individually in an IAC booth. The apparatus and two-interval, forced choice method used have been described previously.

RESULTS

The observed mean attenuation values (protected minus unprotected detection thresholds) are shown as a function of centre frequency for one-third octave noise bands in Figure 1. The parameters are protector type and sex. Across the six groups, attenuation increased from a range of 15 - 27 dB at 250 Hz to 26 - 46 dB at 6300 Hz. The highest attenuation throughout the frequency range was achieved by males using the E-A-R sponge plug and the lowest, by females using the Bilsom Soft. A two-way analysis of variance revealed that both protector and sex were significant main effects. Using the E-A-R plug, males achieved significantly greater attenuation than females (p < .05 or better) at 1000, 3150 and 6300 Hz. Using the Bilsom Soft, males achieved greater attenuation than females only at 6300 Hz. The data for males and females using the Wilson Sound Silencer were not statistically different at any of the frequencies tested.

Table 1 shows a comparison between the mean attenuation and standard deviation specified by the manufacturer and the data obtained in the experiment. Across frequency and protector type the real world values generally fell short of the manufacturer's specification. A match within 5 dB was attained by the Wilson plug at all frequencies for both males and females, and by the E-A-R for males at all frequencies and for females at 3150 Hz and 6300 Hz.

Rank order correlations between the unprotected detection threshold and attenuation were significantly negative in ten of the thirty conditions. No frequency related trend was evident in the data, except that 3150 Hz produced a significant effect in five of the six protectors by sex subgroups.

DISCUSSION

The results described above show that achieved attenuation for particular types of protector will be affected by the size of the ear canal, as exemplified by differences in males and females. Across protector type and frequency differences due to sex ranged from 0 to 10 dB. The smallest difference was observed for the Wilson plug. The latter was available in two sizes, regular and small. Each
subject was fitted with both to determine the most appropriate size. Typically females used the small size, and were able to achieve attenuation values virtually identical to or a few dB better than males.

In order to further test the hypothesis that the real world attenuation achieved for various protectors is related to size and shape of the ear canal, bilateral full canal molds are being made in the same 120 subjects who participated in the experiment. Values for various parameters will be correlated with the attenuation achieved at each frequency within protector type.

REFERENCES


ACKNOWLEDGMENTS

Supported by Ontario Ministry of Labour grant.

Table 1: Attenuation Predicted and Achieved

<table>
<thead>
<tr>
<th></th>
<th>E-A-R</th>
<th>WILLSON</th>
<th>BILSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 Hz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spec.</td>
<td>31.3</td>
<td>(3.3)</td>
<td>23.0</td>
</tr>
<tr>
<td></td>
<td>25.0</td>
<td>(3.2)</td>
<td></td>
</tr>
<tr>
<td>M.</td>
<td>26.9</td>
<td>(5.7)</td>
<td>21.1</td>
</tr>
<tr>
<td></td>
<td>18.9</td>
<td>(9.3)</td>
<td></td>
</tr>
<tr>
<td>F.</td>
<td>24.3</td>
<td>(6.1)</td>
<td>23.2</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>(7.1)</td>
<td></td>
</tr>
<tr>
<td>500 Hz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spec.</td>
<td>34.1</td>
<td>(2.1)</td>
<td>24.0</td>
</tr>
<tr>
<td></td>
<td>29.0</td>
<td>(2.5)</td>
<td></td>
</tr>
<tr>
<td>M.</td>
<td>31.5</td>
<td>(6.0)</td>
<td>25.7</td>
</tr>
<tr>
<td></td>
<td>21.2</td>
<td>(9.6)</td>
<td></td>
</tr>
<tr>
<td>F.</td>
<td>27.9</td>
<td>(6.5)</td>
<td>26.1</td>
</tr>
<tr>
<td></td>
<td>18.0</td>
<td>(8.3)</td>
<td></td>
</tr>
<tr>
<td>1000 Hz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spec.</td>
<td>34.0</td>
<td>(2.3)</td>
<td>26.0</td>
</tr>
<tr>
<td></td>
<td>34.0</td>
<td>(3.1)</td>
<td></td>
</tr>
<tr>
<td>M.</td>
<td>33.6</td>
<td>(0.2)</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>23.9</td>
<td>(9.9)</td>
<td></td>
</tr>
<tr>
<td>F.</td>
<td>26.5</td>
<td>(6.6)</td>
<td>26.6</td>
</tr>
<tr>
<td></td>
<td>19.9</td>
<td>(6.6)</td>
<td></td>
</tr>
<tr>
<td>3150 Hz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spec.</td>
<td>40.8</td>
<td>(1.8)</td>
<td>41.0</td>
</tr>
<tr>
<td></td>
<td>45.0</td>
<td>(3.4)</td>
<td></td>
</tr>
<tr>
<td>M.</td>
<td>41.3</td>
<td>(5.0)</td>
<td>37.8</td>
</tr>
<tr>
<td></td>
<td>33.9</td>
<td>(10.7)</td>
<td></td>
</tr>
<tr>
<td>F.</td>
<td>36.8</td>
<td>(6.9)</td>
<td>38.3</td>
</tr>
<tr>
<td></td>
<td>30.5</td>
<td>(7.5)</td>
<td></td>
</tr>
<tr>
<td>6300 Hz</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spec.</td>
<td>39.9</td>
<td>(2.0)</td>
<td>38.0</td>
</tr>
<tr>
<td></td>
<td>41.0</td>
<td>(4.4)</td>
<td></td>
</tr>
<tr>
<td>M.</td>
<td>45.9</td>
<td>(8.0)</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td>33.8</td>
<td>(12.6)</td>
<td></td>
</tr>
<tr>
<td>F.</td>
<td>36.0</td>
<td>(6.9)</td>
<td>39.3</td>
</tr>
<tr>
<td></td>
<td>26.4</td>
<td>(10.1)</td>
<td></td>
</tr>
</tbody>
</table>

* 3000 Hz
+ 6000 Hz

Fig. 1 Average attenuation achieved with three insert protector types.
EVALUATION OF HEARING PROTECTOR PERFORMANCE IN UK.

P. D. Wheeler, Institute Of Sound And Vibration Research, University Of Southampton, UK.

ISVR have for the past ten years carried out attenuation measurements on hearing protectors to BS 5108:1974, a subjective test of occluded/unoccluded threshold using third-octave bands of noise at centre frequencies from 63 Hz to 8000 Hz. BS 5108 calls for fifteen subjects to be tested twice, on separate occasions, having fitted the protector themselves, under supervision, in broadband noise at a level between 70-75 dB.

Researchers at ISVR were directly involved in the derivation of the Standard (Howell and Martin, 1973) and the ISVR test centre is one of three facilities operating in the UK, along with the National Acoustical Laboratory, and the University of Salford who both carry out test programmes from time to time. In 1983, BSI adopted ISO 4869 as a revision to BS 5108 with the comment in the national foreword "that more subjects and/or replications than the specified minimum of ten tested once are recommended".

Hearing protector tests are undertaken for research purposes (Martin 1977) and as a commercial service to manufacturers and users, and more than 200 models have been assessed, representing a considerable amount of raw data from which certain analyses may be derived. Conclusions from this test work will be discussed later.

In early 1984, under the auspices of the BSI Personal Safety Equipment Committee rather than the Acoustics Committee, BS 6344 Part I "Industrial Hearing Protectors - Ear Muffs" was published. The Standard is concerned primarily with physical performance, durability, and hygiene, but also provides sizing information, and requires that attenuation data according to BS 5108:1983 are provided to users.

The core of the Standard, which is of primary interest to acousticians, is a sequence of durability tests preceeded and followed by insertion loss measurements on an artificial head (following ISO/DIS 6290). Any change in insertion loss following durability testing is required to be less than 4 dB at any of the test frequencies (third-octave bands at 125, 250, 500, 1000, 2000, 3150, 4000, 6300, and 8000 Hz). Although a compromise between statistical requirements and practicality, both before and after insertion loss tests must be repeated, on each cup, until the difference between successive cumulative mean insertion losses is less than 0.5 dB at each test frequency, therefore ensuring that, in practice, any overall change in insertion loss of the order of 4 dB is real rather than due to random variability. Normally, this means that five successive insertion loss measurements are made on each cup before and after testing.

The durability tests comprise three drop tests at room temperature from a height of 2m onto a steel plate, a vibration test at 40 m/sec² at 50 Hz for three hours, and a headband cycling test in which the muffs are subjected to 200 cycles per minute at a rate of 10-12 cycles/minute. In this last test, the headband force must not change by more than 25% after allowing 16 hours recovery time, after which headband force and insertion loss are remeasured.

BS 6344 has been adopted as the basis for a British Standards "Kitemark" Safety Assurance Scheme. Nine manufacturers are licensed to produce some twenty-five models of ear muffs complying with BS 6344, and hearing the "Kitemark" insignia also found on such consumer products as car seat belts, motor cycle helmets, and domestic appliances, will carry a guarantee of comfort, performance and compliance with QA requirements is assessed by type-approval and audit tests carried out by independent laboratories such as ISVR on behalf of BSI. It is hoped that the "Kitemark" will become the basis for UK Government approval of specific protectors for industrial use, thereby placing emphasis equally on product wearability and durability and on selection by attenuation performance.

Further parts of BS 6344 are in draft, dealing with other types of protector such as earplugs, and with use, care and maintenance guidelines.

Although ISO/DIS 6290 has not been confirmed as a Standard, because of its preclusion of the use of a real-time analysis technique and reported variability problems, the UK has accepted its inclusion in BS 6344. The reasons behind this approach are that BS 6344's text was amended to allow real-time analysis, and that the substantial volume of data collected during the drafting of our Standard, showed that good repeatability could be obtained from minute-to-minute and hour-to-hour testing in insertion loss such as is required by BS 6344 before/after scheme, although considerable variability in insertion loss measurements made on single samples of protector from month-to-month had been reported.

The reliability of the measurement technique for periodic quality control and audit testing over many years production remains to be assessed during the life of the scheme.

The manual provides tables of before/after change in insertion loss of 4 dB at any test frequency was initially assessed, in the first drafts, by a single measurement of one cup, which with standard deviations apparently typically in the range 2-4 dB, left something to be desired. (The overall changes in insertion loss of 4 dB effectively yields a standard error of 0.5 dB, thus making changes of 4 dB highly statistically significant, while requiring usually five replications only.)
Of established designs of muffs subjected to BS 6344 testing, some 10-15% have failed either in insertion loss or change in headband force requirements, while others have failed on drop or vibration testing. BS 6344 is now the basis of an International Standard working document (ISO TC94/SC12), where other nations' requirements for durability in environmental extremes are to be accommodated. Tests involving storage at -20°C followed within 20 seconds by three drop tests have yielded spectacular results with some protectors.

Turning to subjective attenuation testing, the question of inter-laboratory variability has been addressed in the literature by Sutton (1982) and Berger (1982). Within the U.K., disparate data from just one type of hearing protector in the early days of testing has been sufficient to cause concern to question the whole basis of laboratory testing of hearing protectors. While commercial confidentiality properly restrains test houses from freely exchanging data, we have been able to compare unselected data in several instances, with favourable conclusions. Figure 1 shows such comparative attenuation data.

Long-term consistency of attenuation data within one test laboratory, where over the space of ten years, some five test officers and numerous changes in subjects have been involved, is illustrated by the data shown in Figure 2 for a formable polymer earplug.

We have examined factors such as non-randomised presentation, and the learning curve for audiometrically-naive subjects, and have concluded that such effects are of little consequence in comparison to the consistently-obtained typical standard deviations of 3-4 dB produced by all three U.K. laboratories using experienced subject panels. However, the attainment of standard deviations of less than 2 dB for products such as disposable earplugs must surely require the use of an extremely refined test methodology and a highly cognisant subject population.

We conclude that in the search for a consistent laboratory technique for the measurement of hearing protector attenuation that is related to the performance that might be readily attained in practice, the U.K. approach has some value, and that on this basis, the statistically-based method of adopting the mean minus one standard deviation attenuation as 'assumed protection' is applicable.

References
MEASUREMENTS OF EAR PROTECTOR ATTENUATION WITH AN ACUSTO-MECHANICAL MODEL OF THE HEARING SYSTEM

C. Giguere, H. Kurov, S. M. Abel and J. Schroeter

Mount Sinai Hospital Research Institute, Room 843, 600 University Ave, Toronto, Ontario, Canada M5G 1X5

and

Institute of Biomedical Engineering, University of Toronto, Toronto, Ontario, Canada M5S 1A4

I. INTRODUCTION

Different methods have been developed to evaluate the acoustic attenuation of hearing protectors. Physiological methods, i.e. microphone-based techniques, seem to provide the best results for the evaluation of hearing protectors [1]. We developed a new head simulator and an artificial ear canal suitable for measurements with earmuffs and earplugs of the KEMAR manikin. The acousto-mechanical model will permit the evaluation of the complex characteristics of hearing protectors under various conditions of noise intensities. In this paper, we present some preliminary experimental attenuation measurements obtained under continuous noise conditions.

II. THE ACUSTO-MECHANICAL MODEL

Our head simulator is based on the commercially available KEMAR manikin since it is a widely accepted device for acoustic research and is amenable to standardization. However, the KEMAR manikin was not designed in the first place for hearing protection measurements and thus, some modifications were necessary. The construction of the new head simulator has been made as modular as possible so that we can refine the different parts separately whenever new needs arise.

In its present stage, the device is composed of the following elements:

1. a head unit in two symmetric left and right halves, cast from a KEMAR original and made of a hard aluminum-filled epoxy;
2. a cylindrical base used to bolt both halves into a single unit;
3. a neck unit which is optionally either (a) rigid, or (b) compliant to simulate the vibration of the head in the sound field; and
4. a pentagonal-shaped cavity in the left side of the head unit, housing one of several insertable ear modules coated with an appropriate mechanical simulation of the circumaural and intracranial human skin.

Three of these ear inserts have been designed. They are all identical in shape, approximating the circumaural contours of the KEMAR head by a three-plane geometry. The exact shape of all inserts has been made cylindrical with a diameter of 7.5 mm. However, the three ear inserts are covered with different thicknesses of an artificial silicone rubber skin. The first ear insert accommodates an artificial circumaural skin of 3 mm thick on the area above the ear and of 6 mm thick on the area below the ear. Furthermore, the ear canal is lined with a 5 mm thick artificial intracranial skin. The second ear insert includes a hard plastic circumaural surface and an artificial ear canal skin lining of 2.5 mm thick. A third ear insert was designed that is made entirely of rigid acrylic plastic and therefore has a hard circumaural surface and ear canal lining. This ear insert is used as a "control insert" to assess the effects of artificial skin on the attenuation of hearing protectors. All other features of the KEMAR manikin have been retained such as the torso, the large pinnae that fill into the ear inserts and the physical dimensions of the head. The sound pressure is measured at the eardrum location of the device using a 1/2" microphone via a Zwiebelki ear simulator.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

Sound Isolation of the Device

The first test carried out was aimed at characterizing the inherent acoustic isolation of the head simulator. This was performed by carefully sealing the ear canal of the control insert with plasticine. The pinnae were not used for this experiment. The acoustic isolation was then obtained by measuring the insertion loss, IL, of the device. The insertion loss is defined as the level difference in dB between the open and unoccluded ear sound pressure, Po, and the closed or occluded ear sound pressure, Pc, i.e.:

\[ IL (dB) = 10 \log_{10} \left( \frac{P_o}{P_c} \right) \]  \hspace{1cm} (1)

The experiment was carried out in a diffuse field inside a large room with hard walls, floor and ceiling. The sound pressure level was measured for 24 1/3-octave bands of noise from 50 to 10000 Hz center frequency. The resulting acoustic isolation of the head simulator was found to be 56 dB or more for all bands. The isolation of the device is therefore greater than the bone conduction threshold [3] for all 1/3-octave bands tested.

Attenuation Measurements of Hearing Protectors

The insertion loss of five earmuffs and of five earplugs was measured over the nine standard 1/3-octave bands (centered at 63, 125, 250, 500, 1000, 2000, 3150, 4000, 6300 and 8000 Hz) according to equation (1). This experiment was carried out to study the effect of the artificial skin on the insertion loss of hearing protectors. The results were also used to compare our objective measurements to the subjective REAT-method. In the real situation, there exists a second path reaching the ear than the direct air-conducted sound. As a result, it was necessary to correct our insertion loss measurements for these sound pathways for a direct comparison with the REAT-method (ASA STD 1-1975). The insertion loss was corrected for the bone-conducted sound and for the physiological masking due to the physiologically noise sources [3] according to:

\[ IL (dB) = 10 \log_{10} \left( \frac{10^{-IL/20} + 10^{-10BC/20}}{10^{-10IL/20} + 10^{-10BC/20}} \right) + \Delta \]  \hspace{1cm} (2)
where IL is the corrected insertion loss in dB. IL is the measured insertion loss in dB from equation (1). BC is the bone-conduction to air-conduction level difference in dB from measurements provided by the E-A-R Corporation and PM is the physiological masking in dB from reference [3]. Without any knowledge of the phase relationship between the air-conduction signal, the bone-conduction signal and the physiological masking, equation (2) assumes they are all in phase. This assumption gives the most conservative estimate of the attenuation.

The measurements with the five earmuffs were performed in the E-A-R Reverberation Room, which complies to ASA STD 1-1975, courtesy of the E-A-R Corporation. The effect of the artificial circumaural skin on the attenuation of earmuffs was studied using two different ear inserts: the artificial circumaural skin ear insert and the hard control insert. The results showed that the mechanical properties of the circumaural surface in contact with the earmuffs affect only slightly their insertion loss. The corrected insertion loss of the earmuffs was then calculated from equation (2). The results for one earmuff is shown in figure (1). The artificial circumaural skin ear insert was used for this experiment. The results are compared to subjective measurements performed at the E A R Corporation on the same device according to ASA STD 1-1975. As can be seen, there is a good agreement between the two methods. The same conclusion was drawn with two other earmuffs for which subjective measurements were available.

The measurements with the five earplugs were conducted at the E-A-R Corporation and at the Mt. Sinai Hospital Research Institute. The effect of the artificial intraaural skin on the attenuation of earplugs was studied using different ear inserts. The results are shown for one earplug in figure (2a) where the insertion loss was measured for three thicknesses of artificial intraaural skin (0.0 mm, 2.5 mm, 5.0 mm). As can be seen, the mechanical properties of the skin lining inside the ear canal have a very great influence on the attenuation of the earplug. Similar results were obtained with the four other earplugs tested. The measurements with the 2.5 mm thick artificial intraaural skin insert were then used to calculate the corrected insertion loss of three earplugs (two rubber plugs and one foam plug) according to equation (2). The results for one rubber earplug are shown in figure (2b) and they are compared to subjective attenuation measurements. As can be seen there is a good agreement between the two methods. A similar agreement was obtained with the other rubber earplug tested. However, the results with the foam earplug were found to be 5 to 10 dB lower than expected by the KEAT method for frequencies below 1000 Hz.

Further work will be aimed at improving the present design for foam earplugs and at measuring the phase characteristics of the insertion loss of different hearing protectors.

[Work conducted under a contract from the Defence and Civil Institute of Environmental Medicine, Canada]

REFERENCES

EVALUATION OF HEARING PROTECTION FOR EARMUFFS IN CLOSED SHOOTING FIELDS

M. Nicoli and M. Cossa
Istituto Superiore di Sanità, Acoustic Dept. Via Regina Elena 299, 00161 Rome, Italy

An experimental investigation for performance evaluation of ear protectors has been carried out, analysing the sound attenuation in the range of 8 different models of earmuffs to be used in closed shooting fields.

The investigation has started from the analysis of sound pressure levels of pistol shots in field conditions. In tab.1 are reported the maximum sound levels related to different time weighing constants (Fast, Impulse, Peak) and measured as Single Event Levels (SEL).

<table>
<thead>
<tr>
<th>WEAPON</th>
<th>MAXIMUM VALUE</th>
<th>SEL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FAST</td>
<td>IMPULSE</td>
</tr>
<tr>
<td></td>
<td>dBA</td>
<td>dBA</td>
</tr>
<tr>
<td>Beretta mod92</td>
<td>124.6</td>
<td>126.2</td>
</tr>
<tr>
<td>Smith &amp; Wesson</td>
<td>125.7</td>
<td>127.5</td>
</tr>
<tr>
<td>M 12</td>
<td>124.0</td>
<td>125.4</td>
</tr>
<tr>
<td>Mean Values</td>
<td>124.8</td>
<td>126.4</td>
</tr>
</tbody>
</table>

Tab.1 Analysis of maximum sound level of pistol shots evaluated with Fast, Impulse and Peak time Weighting and measured as SEL.

Using a Real Time FFT Analyzer it has been also possible to obtain level-frequency spectrum and amplitude-time diagram for a pistol shot (fig. 1 and 2).

Fig.1 Level-frequency spectrum of a pistol shot

For experimental investigation on sound attenuation of each model of earmuffs it has been employed an objective method based on use of an artificial ear inserted in a head mock-up. This method appears more reliable than the subjective criteria based on analysis at threshold of hearing, since the ear performance at high energy level due to a fire-arm shot is different from ear sensitivity referred to network weighing at threshold level.

The analytical method for calculation of earmuff attenuation has been based on the ISO-NIOSH criterion named "long method", because it has an high reliability and an accurate procedure particularly apt for describing earmuff performance in the presence of specific and repetitive noises as fire-arm shots occurring in the same environmental conditions.

In fig.3 is reported an example of frequency analysis of sound level on artificial ear measured with and without ear protector.

Fig.3 Experimental frequency analysis of attenuation curve for an earmuff (sound levels calculated on the basis of ISO-NIOSH "long method")

Carrying out the composition of frequency components of sound level it has been obtained the overall noise pressure level at the ear without and with earmuff protection.

Comparatively it has been evaluated the difference between the experimental measurements and the manufacturer's attenuation factors for the ear protectors investigated. It results from such data that the homologation factors determined by the "short method" procedure as NRR (Noise Reduction Ratings),

Fig.2 Amplitude-time decaying curve of a pistol shot
are rather higher than experimental results, with a
difference comprised between 5 - 11 dB.

In tab. 2 is reported the attenuation calculated with "long method" procedure for the 8 models
examined.

| Earmuff n°1 overall attenuation 23,8 dBA |
| Earmuff n°2 overall attenuation 19.5 dBA |
| Earmuff n°3 overall attenuation 19.4 dBA |
| Earmuff n°4 overall attenuation 18.1 dBA |
| Earmuff n°5 overall attenuation 16.5 dBA |
| Earmuff n°6 overall attenuation 13.6 dBA |
| Earmuff n°7 overall attenuation 13.6 dBA |
| Earmuff n°8 overall attenuation 11.6 dBA |

Tab. 2 Overall attenuation for 8 earmuffs analyzed on the basis of ISO-NIOSH "long method"

Moreover, it has been calculated the maximum number of explosive events compatible with an allowable daily noise exposure of 90 dBA, evaluated as Equivalent Continuous Level.

Integrating n pistol shots measured as Single Event Levels, it is possible to calculate the noise exposure of personal attending the shooting fields through the following formula:

\[ L_{pde} = 10 \log \frac{1}{T} n 10^{-0.1SE} \text{ dBA} \]

where:

- \( L_{pde} \) = acceptable limit of personal daily exposure to noise, corresponding to 90 dBA for a normal 8 hours working time;
- \( T \) = effective daily permanence time for the staff attending the shooting field (evaluated in 4 hours = 14,400 seconds);
- \( SEL \) = Single Event Level of a pistol shot corresponding to 117.4 dBA.

As a consequence:

\[ n_i = \text{antilog} 0.1(901 + E_{A_i} - 10 \log 0.1SEL + 10 \log T) \]

where:

\[ E_{A_i} \] = sound attenuation for each model of the examined earmuffs.

In conclusion it results that the maximum number of explosive events allowable with an acceptance limit of 90 dBA as daily exposure level is:

- for n°1 earmuff: \( n_1 = 6.310 \) shots
- for n°2 earmuff: \( n_2 = 2.345 \) shots
- for n°3 earmuff: \( n_3 = 2.295 \) shots
- for n°4 earmuff: \( n_4 = 1.700 \) shots
- for n°5 earmuff: \( n_5 = 1.180 \) shots
- for n°6 earmuff: \( n_6 = 600 \) shots
- for n°7 earmuff: \( n_7 = 600 \) shots
- for n°8 earmuff: \( n_8 = 380 \) shots

The n°1 earmuff therefore allows an exposition to explosive events more than 16 times than the n°8 earmuff; it is useful also to stress that in no case the shot sound level at the ear protected with all models examined reaches a peak value exceeding the acceptable limit of 140 dBA.

References

PERFORMANCE OF HEARING PROTECTION DEVICES AT LOW FREQUENCIES

V.J. Chwojka and C.L. Oatigu
Department of Applied Mechanics, Ecole Polytechnique of Montreal, P.O. Box 6079, Station A, Montreal H3C 3A7

INTRODUCTION

In spite of a generally satisfying aural protection at high frequencies, the performance of hearing protection devices often falls tremendously at low frequencies. There, an apparent amplification effect, instead of insertion loss, might be observed (1). Furthermore, this was subjectively perceived during several low-frequency field measurements such as those with a new type of low-frequency noise controlling barrier. Several hearing protection devices were put to the efficiency test by a dummy head in order to investigate this deficiency particularly at low frequencies up to 500Hz. Measurements took place in our anechoic room and in the field. The insertion loss data were obtained using a dummy head via the methodology of ANSI S1.19-1974 (2) and ANSI Z24,22-1975 (3). Measurements were performed at various angles of incidence following Burkhard's convention. The main interest was basically oriented at a 0° angle of incidence, simulating a common kind of situations such as an operator looking at a major noise source when placed in its proximity. Results of this study generally show a weak performance and possible noise amplification (instead of attenuation) up to 500Hz depending on the type of measured devices. Furthermore, these results match well with the field experimental results of Goff and Blank (1). Finally, some successful modifications applied to standard Supra-A muff have shown a good possibility of insertion loss improvement at the frequency range between 200 and 500Hz.

METHODODOLOGY

The insertion loss (IL) defined as the difference between the eardrum sound pressure levels with and without the muffs in place, was measured by an anthropometric dummy head, of KEMAR type. It has 1/4" calibrated condenser microphones but has no Dwieloeni coupler to model eardrum impedance (4). However, this apparatus entirely satisfies ANSI S1.19-1974 and has become our arbitrary testing unit. Anyhow, it was not a subject of this study to compare our directional sound field measurements in anechoic conditions to a diffuse sound field measurement in accordance with S1.19-1972. It was supposed that only slight differences could be observed, such as demonstrated in ref.5.

MEASUREMENTS

Two series of measurements were performed independently in our anechoic room and in the field (Fig. 1). The calibrated dummy head was exposed to white noise generated by Klipsch La Scala powered by a Crown PSA-2 amplifier and driven by a B & K 1405 random noise generator. Analysis was performed on a 2 channel FFT analyzer SD-375 and consequently plotted. Six types of muff and 3 types of earplugs were tested to determine their insertion loss characteristics.

RESULTS

Typical characteristics of hearing protection devices measured in laboratory and in the field at 0° azimuth are shown in Fig. 2. The muffs performed generally better (practical low-frequency limit about 400Hz) than the selected earplugs which were almost inefficient up to 800Hz. All the earplugs had a major deficiency in fitting to human auditory canals (temporary subdued fitting), additionally influenced by the jaw motion, subjective geometry and particularities (hairs) in the auditory canal, etc. The earplugs, in general, seemed to fit and stick better to the plastic coated artificial canals, however their low-frequency performance was rather deceiving when compared to that of muffs or Edwards & al. experience (6). On the other hand, the field measurement results, more pessimistic than the laboratory data had a similar trend. Furthermore, those results matched well with Goff and Blank field experimental data (Fig. 3), and confirmed a possibility of modifications (max. 7dB at 250Hz) at low frequencies. Finally, a successful modification (supplementary liquid filled cushion, like acoustic membrane cup cover) applied to a standard Supra-A muff showed a good possibility of the insertion loss improvement at 1/3 octave frequency bands between 160 and 500Hz (Fig.4).

CONCLUSION

In this limited study of hearing protection devices, the main goal was to indicate what might be expected as a degree of muff protection at low frequencies and to verify the Goff and Blank field results. The results of this test indicate that the typical earmuffs do not significantly alter the low-frequency noise reaching the eardrum, neither can they provide any significant degree of hearing protection at low frequencies. They also prove clearly a spurious performance and possible amplification for the tested specimens at low frequencies and confirm the results already presented by Goff and Blank.

REFERENCES


Anechoic room

Fig. 1 Measuring equipment for testing muff performance

Laboratory source: KLIPSCH La Scala
Generator: B & K 1403
Amplifier: Crown PSA-2

Fig. 2 Insertion loss of some hearing protection devices
PELTOR H-7 (Laboratory)
PELTOR H-7 (Field)
SUPRA "A" (Laboratory)
SUPRA "A" (Field)
POLY-CONE LRAL plug (Lab)
P.O.P. Bilsom cotton plug (Lab)

Fig. 3 Comparison of a several field measurements with Coff & Blank field experimental results
Muff type: PELTOR H-7
SUPRA "A"
Average of all 40 muffs (G&B)
Type A, Test 30 (G & B)

Fig. 4 Excess insertion loss of modified vs standard SUPRA "A" muffs
ANAMORPHICAL MEASUREMENTS ON EAR-MUFFS

J. Pifretzchner, A. Moreno
Instituto de Acústica, C.S.I.C.
Serrano 144, 28006 Madrid Spain

INTRODUCTION

In a previous paper [1] a comparative analysis between the insertion loss (I.L.) and transmission loss (T.L.) on ear-muffs (E.M) in a frontal incidence acoustic free field was considered. Experimentation was made employing an acoustical text mixture (A.T.F.), following ISO DIS 6829, and an ensemble of five people with quite different contextures and head sizes. An E-M of fluid filled cushions (Wilsom KU 2301) and pink noise filtered in 1/3 octave bands were among the experimental conditions. It was demonstrated that differences reported in the literature, between I.L. and T.L. are exclusively due to the different diffraction effects present in each case.

The present paper summarizes the results obtained in a quasi diffuse sound field comparatively to the previous ones and to subjective tests (I.S.L) made at the real ear thresholds in both free and diffuse fields (F.P., D.P.).

EXPERIMENTAL ANALYSIS

Experiments were made inside an anechoic chamber. In the case of a frontal F.P. the acoustic source was a little loudspeaker box distant 3 m from the subject or A.T.F. In accordance with BS 5108 the D.P. was produced through four loudspeakers disposed in the corners of a regular tetrahedron and fed by pink noise (bands of 1/3 oct width or broad band). In order to obtain the non coherence of the sources, each loudspeaker was fed by a digital random noise generator tuned with different clock frequencies. In the case of continuous noise bands of 1/3 octave width, a digital delay line was inserted before each power audio amplifier. The non coherence of the signals fed to each loudspeaker was checked through the cross-correlation of the electrical signals. The acoustic field was then tested in the centroid of the tetrahedron by the successive emission of 1, 2 or 4 loudspeakers simultaneously and checking afterwards that the frequency response in all the three cases differed by 3 dB approximately. The symmetry of the test sound fields at the subject’s head position was in accordance with the standards ASA 1971, ISO DIS 68290 and BS 5108.

The acoustic isolation of the A.T.F. fulfilled ISO/DIS 68290 in any test frequency band in the range of interest.

For the subjective measurements in frontal F.P., pure tones were used (ASA 224.22). For the semiadjecti- tive (un people) and objective (un A.T.F) broad band pink noise or filtered in 1/3 octave bands were the test signals both for the F.P. or D.P. Subjective measurements in D.P. with 1/3 bands of pink noise were also carried out. The analysis of the signals was performed through an FFT analyzer (Nicolet 440) and/or an acoumograph (binary electronic device widely used in laboratoris in a large type of acoustic problems). The second instrument has a great frequency discrimination that makes the FFT device impossible to obtain with 1/3 octave FFT device. A pair of miniature electret microphones was chosen from a set of microphones in order to have two different frequency response curves (flat in the range 0.5 - 10 kHz). In the semiadjective measurements (I.L.O, T.L.O) one of these microphones was inserted in a prosthesis, that adapts to the ear canal entrance.

Although the position inside the E.M can be anywhere because the diffuse field created inside it [1], that position is more adequate than any other, i.e. fixed inside the protector, in order to avoid erroneous level measurements due to the resonances of the cups. In addition it is a good reference point to the measurement of the sound pressure level in absence of the E.M.

For T.L. measurements it is necessary to select an external point representative of the total energy incident on the E.M in both cases: free of diffuse fields. This representative point has been proved to exist in F.P. [1].

In order to find this reference point in D.P., one of the cups was divided in 15 equivalent areas. Comparing all the frequency responses obtained in the central points of each area, it was found that the most representative point was situated near the symmetry center of the cup. For this position deviations from the average response were <1 db for 2 and 5 kHz and <0.5 db in the reminder frequencies. It was demonstrated in [1] that for this situation the relationship between I.L. and T.L. is only due to geometrical conditions, corresponding to the incremental scattering measured at the ear canal entrance (with the E.M. absent) at the representative external point (with the E.M. positioned). This relationship indicates the equivalence of both magnitudes.

RESULTS AND CONCLUSIONS

Our principal interest was to know about the nearness among IL measurements (or T.L) made in a subjective way on people and those obtained with the A.T.F. or people acting as passive subjects (semia- djective) in free and diffuse fields.

Figures 1a and 1b show in a continuous line the results of ILS obtained with a selected subjects (5 people, 3 replications) in free and diffuse field. Although the number of subjects tested was little, the spread of results was of the same order of that found by K. Brinkmann [2]: less than 7 db for the frequency range studied. Black dots corresponds to IL measurements on the ATF and white dots to the ILO on human subjects. The standard deviations of ILS were of the order of 1 db except the frequency bands of 3 and 4 kHz that reaches 2.2 and 2.9 db. For the ILO measurements the standard deviations were <4 db excepting 5 and 10 kHz with 4.6 and 4.5 db respectively. Analogous values of the standard deviation were found for D.P.

In both figures it can be appreciated the influence of physiological noise (P.N) due to the occlusion effect by the E.M. We have found higher values of P.N in D.P than in F.P. due, any doubt, to the worse discrimination of threshold for 1/3 octave bands of noise than for pure tones. As it is usual in the literature the concordance between ILS and ILO is fairly high up to 2 kHz; below this frequency discrepancies should be due to the bone conduction (B.C.). It is interesting to emphasize that discrepancies found in the high frequency range between F.P. and D.P. curves, give pessimistic attenuation values for frontal F.P. Incidence.

Figure 2 summarizes the differences found as a function of the measuring conditions. ILISO and TISO mean respectively A.T.F. insertion loss in F.P. or D.P. (average of 8 measurements); similarly ILOE and ILOO mean semiajective I.L. on people (average of 6 and 11 measurements) and ILF, ILOE corresponds to subjective measurements (15 measurements). In this way Fig. 2 shows differences between IL results on A.T.F. in both fields. So reduced differences seem due to the simple geometry of the system. The values of the differential scattering measured in the center
of the lateral sides of the ATP in the two cases are
different above 2 kHz giving an absolute limit < 2.5
dB; it occurs similarly with diffraction in the
representative point outside the E-M.

In the measurements on humans, differences above
6 kHz are remarkable, the E-M presents higher iso-
lation in D.F. than in P.F. both in the semiobjective
(2b) and in subjective measurements (2c). The same
tendency is obtained when we consider incremental dif-
frations in each frequency band and in each situa-
tion.

Taking ILS as a reference, ILQ constitutes an
approach better than the others as can be seen in
Figs. 2d, e, f and g. Deviations lower than 3 dB for
that approach are found except for 8 kHz where 4.8 dB
are attained and for frequencies below 500 Hz af-
tected by the P.N.

The comparison between human subjects and A.T.F.
shows the influence of the pinna at 2 kHz not only in
frontal F.R. in accordance with Schröter and Pössel
[3] but also in D.F. The pinna decreases the values of
IL of circumaural hearing protector devices at
mid-frequencies.

The spectrum of differences between subjective
and objective attenuation measurements (2d) have
similar patterns that those exposed by Chilley [4]
except in the 4-10 kHz range where we have found
lower values. This fact is perhaps due to the dif-
f erent cushions incorporated in our E-M (Liquid
filled instead foam filled) and to the nature of the
D.F. or A.T.F very critical at those frequencies.

The difference curves between subjective and
semiobjective measurements in P.F or D.F (figs. 2c
and 2g), shows the same trends that those exposed by

The concordance between subjective measurements
and the ISO ATP is not very high in P.R. and very low
in F.R. It is due not only to the different diffra-
cion and occlusion effects but also by the fact that
the isolation associated to the cups, are function of
the type of incident field (phenomenon well known in
building acoustic).

The best accordance between objective and semi-
objective measurements, on the same human subjects,
have been obtained specially in the case of D.F.
measurements, where differences (excepting the P.N.
range) are of the order of ±3 dB.

It can be conclude that the best test fixture for
objective measurements in ear-muffs is the "real"
human subject equipped with miniature electret micro-
phones placed in the ear canal under D.F. conditions.
These semiobjective measurements are very fast and
permits a great variety of subjects to be tested,
introducing most human characteristics. Hence a
very realistic results can be expected.

REFERENCES

Loss versus Transmission Loss on Ear-Protectors"
5th PASE Symposium, 1985.

tainties of sound attenuation measurements on
hearing protectors". Inter-Noise 85. 1287-1290.

measuring the insertion loss of hearing protecti-
ve devices including simulation of the bone con-

tive measurement of circumaural hearing protector
attenuation", in Personal hearing protection in
industry, edited by P.W. Alberti (Raven, New

[5] A. Martin, "How realistic are standard subjective
 test methods for evaluating hearing protector at-
tenuation", in Personal hearing protection in
industry, edited by P.W. Alberti (Raven, New

This work was partially supported by the CAICT
SOURCES OF VARIABILITY IN NORMAL HEARING SENSITIVITY

D.W. Robinson
Institute of Sound and Vibration Research,
The University, Southampton S09 5NH, England.

The calibration of pure-tone audiometers rests on a standardized measure of central tendency of the hearing threshold levels (HTL) of young optometrically normal persons (ISO 389). Until the recent publication of a complementary standard (ISO 7029) dealing with the effect of age, no account was taken of the dispersion of individual values, which can amount to as much as ±15 dB amongst young people - still more with advancing years. Thus, irrespective of measurement accuracy, an uncertainty attaches to the interpretation of an individual audiogram: a measured HTL of 0 dB may betoken an unimpaired ear of an individual, or it may reflect a loss of hearing in one whose original condition was more sensitive by 10 or 15 dB. Although less obvious, similar uncertainty applies to HTLs of any magnitude. Such uncertainty is usually regarded as an ineradicable limitation of audiology and attributed to inherent biological variability - not a very satisfactory explanation.

The present study shows that the dispersion, and hence the HTL itself, can be decomposed into a number of identifiable parts, to different aspects of the auditory process which can be separately tested. In this way a better picture may be gained of the state of functioning of a given ear.

TEST METHOD

The tests described were performed on a group (0) of 50 older subjects (age 45-63 yr) and a group (1) of 81 young subjects (age 16-27 yr), all having been screened otoacoustically and by extensive questioning for adverse indications in their otological and noise exposure histories; both ears were tested.

1. Pure-tone audiometry was carried out from 0.25 to 8 kHz with a self-recording instrument. Here only the HTLs at 4 kHz, denoted by H, are considered.

2. Brief-tone audiometry was carried out, also by self-recording, using 4 kHz tone bursts with duration t, ranging from 1.5 to 400 ms. Results for 1.5≤t≤40 ms were expressed as linear relations of threshold shift (relative to the 400 ms burst) upon $\log t$, yielding a slope of $k$ dB/decade and an intercept $T$ (an auditory integration time). The parameter of interest is $u = k \log (T/T_0)$; the derivation of the normalizing duration is described below.

3. Octave masking was carried out, again by self-recording, using a 2 kHz (slightly mistuned) tone as masker and a pulsed 4 kHz probe tone. The results were expressed as 4 kHz threshold shifts (relative to H) against masker level. These data were fitted by a linear relation for threshold shifts of 10 dB and upwards, with a slope of $v$ dB/dB as the parameter of interest.

RESULTS

Figures 1, 2 and 3 show the distributions of the relevant variables, H, u, v, for the two groups, on normal probability coordinates oriented so that better hearing is to the left in each case. Note that the abscissa scales differ. Standard deviations are indicated on the curves. The three variables turned out to be highly correlated ($r < 0.001$ in all cases) and only k,v, group 0, was slightly less significant ($r < 0.01$) (see Table 1).

<table>
<thead>
<tr>
<th>Group 0</th>
<th>Group 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>k</td>
</tr>
<tr>
<td>k</td>
<td>-0.68</td>
</tr>
<tr>
<td>u</td>
<td>-0.74</td>
</tr>
<tr>
<td>v</td>
<td>-0.70</td>
</tr>
</tbody>
</table>

These correlations imply that individuals with sensitive hearing (low H) tended to be those with high values of u and v. Moreover, referring to the Figures, retention of hearing sensitivity with age seems to be accompanied by retention of high values of u and v; at the other end of the distributions, less sensitive hearing is associated with more rapidly declining values of u and v.

ANALYSIS

The results are considered in terms of the variance of quantities of the type:

$$ J = H + u + b \cdot k + u \cdot v \quad \ldots \ldots \ldots \ldots \ldots \ldots (1) $$
with $a = k \log T$. The coefficients a, b and c were determined by step-wise multiple regression. Noting that b may be written as log $10^b$, the expression takes the form:

$$ J = H + a \cdot u + c \cdot v \quad \ldots \ldots \ldots \ldots \ldots \ldots (2) $$
With u = $k \log (T/T_0)$ and $T_0 = 10^{-b/a}$. With the coefficients allowed to float without constraint, the var(j) for the O- and Y-groups reduced respectively to 36.4% and 70.0% of the crude var(H). In other words, almost 2/3 of the variance of HTL at 4 kHz is accounted for by the temporal integration and frequency selectivity properties of the cochlea in the older group, and nearly 1/3 in the younger persons. Whereas the crude var(H) differed greatly between the two groups (90.0 dB² compared with 37.7), the residual variance not accountable to the identified sources was more nearly equal (32.7 dB² against 24.6).

In the unconstrained analysis, the coefficient a assumed different values, 1.58 and 1.15 for the O- and Y-groups respectively. These values do not, however, necessarily reflect the underlying auditory process because the 'independent' variables are, in reality, not independent but correlated. It was found that by applying the constraint $a = 1$, the ratio var(j)/var(H) remained at near-optimum values of 31.4% and 70.2% for the two groups. This constraint admits of a simple physical interpretation, since u is just the threshold shift for a pulse of very short duration $T_0$ relative to the conventional (integration-complete) audiometric threshold. The associated values of $T_0$ in this case turned out to be 0.99 ms for the O-group and 3.86 ms for the Y-group. Given that the minimum of var(j) is not a very critical function of $T_0$, particularly in the case of the Y-group, these estimates may not be inherently inconsistent and the lower value is probably the more reliable. $T_0$ may represent some kind of trigger time required to initiate the integration process.

The coefficients c also differed somewhat between the two groups (3.53 against 2.83 for the O- and Y-groups respectively, each with the constraint $a = 1$); again the first of these estimates is probably the more reliable.

Figure 4 illustrates the distributions of $J$ for both groups, to the same scale and with the same conventions as Figures 1. They were calculated with the constraint $a = 1$. Note that the distributions are closely alike except for a horizontal displacement.
INTERPRETATION

By rewriting eqn (2), with $a = 1$, in the form:

$$H = J - u - c_v$$

... (1)

an NTL, as conventionally determined with stimuli having durations of some hundreds of milliseconds, appears as the composite of several components. The first of these, $J$, may be interpreted simplistically as a conductive efficiency measure, influenced by variations in the mechanics of the middle ear. As such it may be subdivisible into its own components (but not by the tests described here). The standard deviation of $J$ at 4 kHz appears to be about 5 dB in a young normal population and to increase only slightly with increasing age, although the mean value changes substantially (by about 20 dB between the $Y$- and $O$-groups in the present tests).

Fig.1: Distribution of $H$
(NHL at 4 kHz)

Fig.2: Distribution of $u$
(temporal integration component of $H$)

Fig.3: Distribution of $v$
(coefficient of the frequency selectivity component of $H$)

Fig.4: Distribution of $J$
(presumed conductive efficiency component of $H$)

The second component, $u$, consists of a gain due to temporal integration of the signal, presumably in the neural rather than the hydrodynamical domain. This gain is characterized by increasing dispersion as the age increases.

The third component, $c_v$, is a further gain due to the mechanism of frequency selectivity; this operates, however, in a less obvious way than the integration gain. A physical interpretation of the term $c_v$ may perhaps be found in the spread of mechanical excitation along the basilar membrane of the cochlea.

The three components were identified both in a young otologically normal population and in an older population selected by the same 'normality' criteria. The effects of the two 'gain' terms, attributed to cochlear function, were considerably greater in the older group, accounting for nearly 2/3 of the variance of NTL at 4 kHz (against about 1/3 in the case of the younger group). The remainder of the variance is presumably attributable mainly to middle-ear conduction, and this component increased only slightly with age as might be expected.

It is suggested that more insight into an individual's auditory status (and possibly the locus of any abnormality) may be gained by applying tests on the lines described here, as an adjunct to conventional pure-tone audiometry, and by resolving the measured NTL into its component parts. The requisite psychosensory tasks present no particular difficulties; the brief-tone audiometry closely resembles ordinary audiometry from the subject's point of view and the tone-on-tone masked threshold test is, if anything, easier to perform.

ACKNOWLEDGEMENTS

The experimental work was carried out by Ben W. Lawton, ISVR Research Fellow, to whom the author is also greatly indebted for many helpful discussions in the preparation of this paper. The results were obtained in the course of a larger study sponsored by the UK Ministry of Defence whose support is gratefully acknowledged.

REFERENCES


A Determination for Hearing Threshold of Bone Conduction by Using Impedance Head

Zhang Juncui, Shen Yang and Shi Shuming

Acoustical Laboratory, National Institute of Metrology, Beijing, China

The hearing threshold values of bone conduction obtained by bone conductors, such as Type B74 could be influenced by the properties of artificial mastoid during its calibrating. We used the method of direct determination for bone conduction a hearing threshold with an Impedance Head in the development of national standard experimented at 1978 to 1980. The results show essentially consistence with ISO Standard data and followed with clinical check in 1981 and 1982.

The Impedance Head we used is from BAK and Type is 8000 with Serial No.344203. Long term examination shows its rather high stability as follows.

Table 1 - The sensitivities of Impedance Head Type 8000 (No.344203) calibrated during 1971 to 1984

<table>
<thead>
<tr>
<th>Voltage sensit. (mV/g)</th>
<th>Charge sensit. (pC/g)</th>
<th>Department for calibrating</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>31.5</td>
<td>35.1</td>
<td>BAK</td>
<td>Nov.5 1971</td>
</tr>
<tr>
<td>30.9</td>
<td>34.6</td>
<td>Beijing City Inst. of Metro.</td>
<td>Aug.1 1979</td>
</tr>
<tr>
<td>30.5</td>
<td>34.5</td>
<td>Mechanical Division of NIM</td>
<td>May 13 1983</td>
</tr>
<tr>
<td>31.3</td>
<td>34.9</td>
<td>Vibration Laboratory of NIM</td>
<td>Aug.10 1984</td>
</tr>
</tbody>
</table>

Owing to the Impedance Head has the facility of delivering a force or an acceleration signal at the test point, the vibration signal generated from a Mini-Shaker (3 8x3505) transferring the subject through Impedance Head will be simultaneously measured to determine the hearing threshold level of bone conduction. The measuring equipments are illustrated in Figure 1.

During experimenting, the Frequency Generator (BPK1022) drive the Mini-Shaker BAK4950, and the force signals from Impedance Head are fed, via a Charge Amplifier BAK2651 to the Measuring Amplifier BAK2606, which used with Band Pass Filter BAK1614. The force values of hearing threshold become the direct readings from the 2606.

The proper static load is achieved by hanging the 4910 and 6000 at 5.4A monitored with Spring Balance made in Germany. Flat test part of the testing position should found and the Impedance Head placed in such a way that there is a contact with the skin around its bottom centre. A special potentiometer is used by operator to adjust the output of 1022, and cut off its signal when pressing the Generator Stop Button. The response of subject is equipped with a small hand held signal light. At each frequency, the subject is at first given a fairly strong signal so that he can identify the frequency he is listening for. Thus the signal level is from zero to just have been heard and once again the signal level is reduced until the subject can no longer hear it, and then brought back his threshold again. Therefore, the subject ought to make 50% error at this threshold point.

The measurements were carried out in an anechoic chamber with ambient noise 14dBa to 16dBa. The spectrum of which is given in Table 2.

Table 2 - Ambient noise spectrum

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>SPL (dB)</th>
<th>Recommen. of ISO 6553</th>
</tr>
</thead>
<tbody>
<tr>
<td>31.5</td>
<td>23</td>
<td>55</td>
</tr>
<tr>
<td>40</td>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>50</td>
<td>12</td>
<td>41</td>
</tr>
<tr>
<td>63</td>
<td>10</td>
<td>35</td>
</tr>
<tr>
<td>80</td>
<td>10</td>
<td>30</td>
</tr>
<tr>
<td>100</td>
<td>10</td>
<td>25</td>
</tr>
<tr>
<td>125</td>
<td>9</td>
<td>20</td>
</tr>
<tr>
<td>160</td>
<td>9</td>
<td>17</td>
</tr>
<tr>
<td>200</td>
<td>9</td>
<td>15</td>
</tr>
<tr>
<td>250</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>315</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>400</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>630</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>800</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>1000</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>1250</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>1500</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>2000</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>2500</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>3150</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>4000</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5000</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>6300</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>8000</td>
<td>1</td>
<td>15</td>
</tr>
</tbody>
</table>

An earphone Type TM-39 (N-41/AR) was used to produce a 40dB hearing level of 1/3 Octave band noise for the masking of non-test side ear during experimenting. 54dB to 74dB differences between with and without masking indicate that masking is necessary. In order to attenuate the airborne sound radiated by the Impedance Head, an earplug

---

Fig. 1 - Measuring equipments of bone conduction hearing threshold

---
Type EAR was used at the same time to insert the entrance of the side ear at high frequencies from 20000Hz. The effectiveness of the experiment is given in Table 3.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Bone conduction hearing threshold with/without EAR earplug</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With EAR</td>
</tr>
<tr>
<td>250</td>
<td>64.5</td>
</tr>
<tr>
<td>500</td>
<td>54.5</td>
</tr>
<tr>
<td>1000</td>
<td>42.0</td>
</tr>
<tr>
<td>1500</td>
<td>35.0</td>
</tr>
<tr>
<td>2000</td>
<td>30.5</td>
</tr>
<tr>
<td>3000</td>
<td>31.0</td>
</tr>
<tr>
<td>4000</td>
<td>34.0</td>
</tr>
<tr>
<td>6000</td>
<td>30.5</td>
</tr>
<tr>
<td>8000</td>
<td>32.0</td>
</tr>
</tbody>
</table>

Table 3 - The effectiveness on bone conduction hearing threshold when test side ear with or without EAR earplug.

106 otological normal Chinese subjects with both sexes aged from 18 to 30 years so that 212 mastoid and 106 forehead positions were under tested. All of them were screened free from the sign of ear disease, had no history of undue exposure to noise or shock, and had been cleaned ear canal for excess wax. Their hearing levels with 40 audiometry were not to exceed 15db at frequency range 125Hz to 6000 Hz. The results of bone conduction hearing threshold on human mastoid taken from the average of 106 subjects are given in Table 4.

Table 4 - Data of bone conduction hearing threshold level on 212 mastoid positions of 106 subjects comparing with others.

<table>
<thead>
<tr>
<th>Pres.</th>
<th>Bone conduction hearing threshold level (Re: 1 μV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>82.5</td>
</tr>
<tr>
<td>160</td>
<td>77.5</td>
</tr>
<tr>
<td>200</td>
<td>72.5</td>
</tr>
<tr>
<td>250</td>
<td>67.0</td>
</tr>
<tr>
<td>315</td>
<td>64.0</td>
</tr>
<tr>
<td>400</td>
<td>61.0</td>
</tr>
<tr>
<td>500</td>
<td>59.0</td>
</tr>
<tr>
<td>630</td>
<td>52.5</td>
</tr>
<tr>
<td>750</td>
<td>48.5</td>
</tr>
<tr>
<td>800</td>
<td>47.0</td>
</tr>
<tr>
<td>1000</td>
<td>42.5</td>
</tr>
<tr>
<td>1250</td>
<td>39.0</td>
</tr>
<tr>
<td>1500</td>
<td>36.0</td>
</tr>
<tr>
<td>1600</td>
<td>35.5</td>
</tr>
<tr>
<td>2000</td>
<td>31.0</td>
</tr>
<tr>
<td>2500</td>
<td>29.5</td>
</tr>
<tr>
<td>3000</td>
<td>30.5</td>
</tr>
<tr>
<td>3150</td>
<td>31.0</td>
</tr>
<tr>
<td>4000</td>
<td>35.5</td>
</tr>
<tr>
<td>5000</td>
<td>40.5</td>
</tr>
<tr>
<td>6000</td>
<td>40.5</td>
</tr>
<tr>
<td>6300</td>
<td>40.0</td>
</tr>
<tr>
<td>7000</td>
<td>-</td>
</tr>
<tr>
<td>8000</td>
<td>40.0</td>
</tr>
</tbody>
</table>

Table 5 - Data of clinical check with the results of bone conduction hearing threshold level.

**Values for these frequencies are derived from the results in one country only.

Because of most audiometric rooms are still working with poor acoustical condition over vast China, we had applied the resultant values in some audiometers as the clinical check at two places, one in Shanen City, Fujian Province with noise level 31dbA to 40dbA representing the ambient noise condition in audiometric rooms located in factories; another in Yangzhou City, Jiangsu Province with noise level 21dbA to 30dbA representing the ambient noise condition in audiometric rooms of hospitals. The number of subjects for both places are all one hundred. 400 mastoid positions had been experienced to test. These clinical check results are shown in Table 5.

Table 5 - Data of clinical check with the results of bone conduction hearing threshold level.

<table>
<thead>
<tr>
<th>Pres.</th>
<th>Hearing level, db</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanen, 1981</td>
<td>21dbA - 30dbA</td>
<td>21dbA - 30dbA</td>
</tr>
<tr>
<td>250</td>
<td>9.4</td>
<td>5.6</td>
</tr>
<tr>
<td>500</td>
<td>6.2</td>
<td>6.2</td>
</tr>
<tr>
<td>1000</td>
<td>6.7</td>
<td>5.0</td>
</tr>
<tr>
<td>1500</td>
<td>6.7</td>
<td>5.9</td>
</tr>
<tr>
<td>2000</td>
<td>4.6</td>
<td>6.3</td>
</tr>
<tr>
<td>3000</td>
<td>6.5</td>
<td>6.6</td>
</tr>
<tr>
<td>4000</td>
<td>4.3</td>
<td>5.8</td>
</tr>
</tbody>
</table>

Conclusions

The results of this experiment could derived several conclusions outline below:

a) The obtained bone conduction hearing threshold level on mastoid locations shows essentially consistence with ISO 7566 data at the frequency range 250Hz to 4000Hz which is used in most of the commercial audiometers.

b) The hearing levels for hearing normal persons obtained at clinical check are no more than 10db. It means that no clinical problem in practice will happen either our data or ISO data. Therefore we have decided to accepted International Standard ISO 7566 as our National Standard.

c) On the data at frequency ranges lower than 250Hz and higher than 4000Hz, it seems some differences exist between ISO data and ours. For this, we hope there will be any international co-operation to experiment with same method and same equipments at several countries, and maybe it is a task on ISO in future.

Bibliography

Simulated Hearing Loss: A Tool for Comparing Normal and Impaired Hearing

Mary Florentine* and Saren Buus**

*Communication Research Laboratory (133 FR) and
**Auditory Perception Laboratory (413 MU), Northeastern University, Boston, MA 02115 USA

Several abnormalities observed in cochlear impairments can be reproduced in a normal ear by masking. When the level and spectral shape of a noise masker is chosen to produce thresholds in a normal listener similar to those of an impaired listener, we call the masked normal ear a simulated impairment. Masking produces the rapid growth of loudness and the reduced dynamic range typical of cochlear impairments (Steinberg and Gardner, 1937). On the other hand, masking appears not to affect frequency selectivity (e.g., Green, Shelton, Picardi, and Gutierrez, 1974), temporal summation (Zwicker and Noll, 1958), or temporal resolution (Florentine and Buus, 1984; Buus and Florentine, 1985). Therefore, it is likely that comparing the performance of real and simulated impairments can help separate the effects of an impaired listener's abnormal intensity perception from the effects of possible alterations in frequency selectivity and/or temporal resolution.

Since 1978, we have been using simulated impairments as a standard control in all our experiments that have studied cochlear impairments. Using this technique we have discovered considerable individual differences in the spectral and temporal processing of cochlearly impaired listeners with very similar audiograms. The purpose of this paper is to report briefly some data on spectral and temporal processing in simulated impairments. In this paper, psychoacoustical tuning curves were obtained to assess spectral processing and gap detection thresholds to assess temporal processing.

METHOD

Spectral Processing

Method

Psychoacoustical tuning curves show a listener's ability to hear one sound in the presence of another by measuring, as a function of frequency, the level of a masker necessary to just mask a soft pulsing tone, the probe. Such measures were obtained in listeners with normal hearing, simulated impairments, and cochlear impairments. Tuning curves were measured monaurally using an adaptive procedure in a two-interval, two-alternative forced-choice paradigm. First, a threshold for the probe presented in quiet was measured. Next, the probe was set 10 dB above threshold and the level of a 50-Hz wide band of noise necessary to mask the probe was measured for each of six center frequencies, to obtain three tuning curves for each listener and probe frequency.

The masker frequencies were 0.43, 0.78, 0.92, 1.08, 1.23, and 1.48 times the probe frequency. The masker was pulsed to mark the observation intervals; it had a duration of 1.300 ms and an interstimulus interval of 250 ms. The probe was a double pulse consisting of two 350-ms tone bursts separated by a 250-ms interstimulus interval. The onset of the first tone pulse was 250 ms after the beginning of the observation interval. The rise-fall times were 20 ms for both masker and probe. The double burst was presented with equal priori probability in either the first or the second observation interval. (For further details of the procedure, see Buus, Florentine, and Mason (1986).)

Results

Figure 1 shows average tuning curves at 1 kHz from three normal listeners (left panel). The tuning curve at the bottom was obtained in quiet and the tuning curve at the top was obtained in the presence of a broad-band masking noise shaped to yield masked thresholds within 3 dB of the quiet thresholds of impaired listener TG (described below). At the two lowest frequencies of the tuning curve, the level difference between the masker and the probe is larger for the unmasked than for the masked tuning curve; otherwise, the tuning curves appear reasonably similar. The tuning curve at the right was obtained from an impaired listener, TG, with a bilaterally symmetrical loss of predominantly cochlear origin. [For details of his audiometric profile, see Buus, Schaff, and Florentine (1984, p. 78).] The hearing loss in his right ear averages about 40 dB HL through 2 kHz and gradually decreases to 60 dB HL at 8 kHz. A comparison of the tuning curves for TG and the unmasked normal listeners could be interpreted as TG having somewhat impaired frequency selectivity. However, a comparison with the average tuning curves obtained in the simulated losses indicates that TG's frequency selectivity does not differ from that of normal listeners when the effect of elevated thresholds, reduced dynamic range, and recruitment are taken into account.

Temporal Processing

Method

This experiment shows the ability of listeners with normal hearing, simulated impairments, and cochlear impairments to hear a silent pause, a gap, in a continuous noise. The minimum detectable gap duration, MDG, in a low-pass cut-off of 7 kHz noise was measured monaurally as a function of SPL in 5- or 10-dB steps. An adaptive procedure with a two-interval two-alternative forced-choice paradigm with feedback was used. The gap occurred in one of two intervals, which were marked by lights, with equal priori probability, and the listener's task was to judge which interval contained the gap. The gap was produced by turning off and on the noise with fall and rise times of 1 ms. (For details of this procedure, see Florentine and Buus, 1984.)
Results

Figure 3 shows the MDGs as a function of level for six normal listeners, two listeners with simulated impairments, and two impaired listeners. The shaded area shows the range of plus and minus one standard deviation around the mean for the normal listeners. Their MDGs decrease from 25 ms at an overall level of 20 dB SPL to about 5 ms at 60 dB and above. As indicated by the small range, individual listeners’ MDGs differ relatively little from the mean, except at low levels. The open circles show the MDGs for PM who is 22 years old. He has a mildly sloping moderate high-frequency hearing impairment of predominantly cochlear origin. His MDGs are elevated near threshold, but decrease rapidly with increasing level. Above 60 dB SPL, they are identical to the MDGs for normal listeners. The asterisks show the MDGs for two normal listeners with simulations of PM’s impairment. Because their individual results are very similar, only the means are shown. The MDGs for the simulated impairments do not differ substantially from those for the real impairment. Therefore, it seems that PM has normal temporal resolution and the elevation of his MDGs near threshold results from the decreased sensation level of the noise. The open triangles show the results for RT who is 20 years old. She has a hearing impairment of predominantly cochlear origin and her thresholds differ no more than 15 dB from PM’s at any frequency between 125 and 18,000 Hz. For PM and RT’s audiometric profiles, see Buus et al. (1984, p. 78) and Florentine and Buus (1984, p. 453). Despite the similarity in age and audiograms, RT’s MDGs are different from PM’s. As expected, her MDGs are elevated near threshold and decrease rapidly with increasing level. In contrast to PM’s MDGs, however, RT’s MDGs are elevated at high levels. Between 80 and 90 dB SPL, her MDGs are 1.5 to 2 times greater than those for PM’s real and simulated impairment. Therefore, it seems that RT has reduced temporal resolution, but most of the elevation of her MDGs at low and moderate levels probably results from the reduced SL of the noise. The comparison of normal and impaired listeners could be interpreted as both impaired listeners having reduced temporal resolution. However, the comparison of simulated and real impairments clearly shows that only one of the impaired listeners truly has reduced temporal resolution. Similar results have been obtained for other impaired listeners (see Florentine and Buus (1984)).

Discussion

These examples of comparisons between normal hearing, simulated impairments, and real impairments clearly demonstrate the diversity of results that are obtained in impaired listeners. Comparisons between cochlearly impaired and normal listeners are most relevant for comparison of the range of the 1 SD around the average MDGs for six normal listeners is shown by the shaded area. All points are at multiples of 5 dB. For clarity, points have been shifted 1.5 dB to the left for KL + MA and 1.5 dB to the right for PM.

Acknowledgments

We thank Chris Mason for helpful comments. This work is supported, in part, by NIH-NINDS grants ROINS18280 and RR07143.

REFERENCES


THE EXTRA EFFECT OF MASKER FLUCTUATIONS ON THE SRT FOR HEARING-IMPAIRED LISTENERS

J.M. Pesten and R. Plomp

Experimental Audiology, ENT Department, Free University Hospital, de Boelelaan 1117, 1081 HV Amsterdam, The Netherlands

The speech-reception threshold (SRT) in noise for hearing-impaired listeners has been studied extensively in the past decade and was described quantitatively by Plomp (1978) in a simple model containing two parameters related to hearing loss: 'A' for attenuation and 'D' for distortion. Hearing loss for speech (SHL) in noise is presented by the parameter D and hearing loss for speech in quiet by the sum of A and D, both in decibels. On the basis of several studies, Plomp (1978) concluded that for large groups of hearing-impaired listeners SHL in noise increases, on the average, with 1 dB for every 3 dB increase in the SHL in quiet. This result was well understood in terms of the hearing loss at which people start to complain about audibility handicap.

More recently, Duquesnoy (1983b) and Plomp (1986) published for various groups of listeners (elderly, listeners with noise-induced hearing loss, and listeners with ear pathology) considerably lower values for D than those presented in some of the earlier studies. A possible explanation for this discrepancy is, that some of the data used by Plomp (1978) were not obtained with a steady-state noise masker but, for instance, with competing speech as a masker (Carhart and Tillman, 1970) or with a fluctuating noise (Kell et al., 1971). Under such conditions rather large differences in masked threshold were found between impaired and normal-hearing listeners. Also Duquesnoy (1983a) found for a group of elderly listeners considerably larger SHL values with competing speech than with noise as a masker.

In this study we tried to bridge the threshold differences between continuous noise and speech as a masker by measuring also the speech-reception threshold (SRT) for sentences in modulated noise as a function of modulation frequency. Speech, however, is not only intensity-modulated but it is also characterised by fluctuations in the frequency domain, like the changes caused by shifting formants. In order to include this factor and to estimate its contribution to the masking of speech, we measured also the SRT for sentences masked by a noise stimulus split in two frequency bands (one below and one above 1000 Hz) which two bands were modulated in anti-phase.

METHOD

Listeners were asked to reproduce short meaningful Dutch sentences, recorded of a female speaker, and presented against a noise background with a spectrum equal to the long-term average spectrum of the sentences. In the modulated-masker conditions the noise was sinusoidally intensity modulated (100%) with modulation frequencies of 4, 8, 16, and 32 Hz. Each of the four modulation rates was tested twice: once for in-phase modulation over the whole spectrum and once for frequencies beyond 1000 Hz modulated in anti-phase. Additionally, nonmodulated noise and continuous discourse were used as maskers, making a total of ten measuring conditions. The discourse masker, read by the same female speaker who read the signal (equal long-term average spectra), was made unintelligible by presenting it time-reversed. The sequence of measuring conditions was counterbalanced for every ten subjects according to a diagram-balanced Latin square.

In all conditions the masker was presented at a sound-pressure level of 75 or 80 dBA for the normal-hearing and hearing-impaired listeners, respectively. Each SRT (50% correct) was determined with a list of 13 sentences, unknown to the listener, by using a simple up-and-down procedure for the presentation level. The procedure requires from the listener correct replication of the entire sentence for a correct response. Each condition, with a different list, started with the first sentence below threshold. This sentence was repeated until the subject was able to reproduce it correctly; with each repetition, the level of the sentence was raised by 4 dB. All following sentences were presented only once and the step size used was -2 dB after a correct response and +2 dB after an incorrect response. The average presentation level after the fourth sentence is taken as the SRT for that particular condition.

Twenty university students volunteered as normal-hearing listeners in this study and were tested monaurally at the ear of their preference (75 dBA masker level). The hearing-impaired listeners were twelve pupils of a highschool for the hearing-impaired. They had sensorineural hearing losses and were tested at their better ear (80 dBA masker level). Pure-tone acuity (PTA, average hearing threshold for 300, 1000, and 2000 Hz) ranged from 12 to 48 dB in this group. Additionally, three listeners with a more severe loss (average PTA is 57 dB) were tested for a masker level of 90 dBA. Because the number of hearing-impaired listeners is not a multiple of ten, the measuring conditions were not completely counterbalanced for this group.

RESULTS

The results for the two groups of subjects are given in fig. 1. The standard deviation between subjects, averaged over all conditions, is 2.3 dB for the hearing-impaired listeners and 1.6 dB for the normal-hearing listeners. Only with speech as a masker much larger interindividual differences are found for the normal-hearing listeners ($\sigma = 5.1$ dB).

Fig. 1: Average speech-reception threshold in modulated noise as a function of modulation frequency for 20 normal-hearing listeners (open symbols) and 12 hearing-impaired listeners (closed symbols). The circles are for in-phase modulation of the whole spectrum and the squares are for conditions in which the frequencies below and above 1000 Hz are modulated in anti-phase.
For optimum modulation of the noise normal-hearing listeners gain nearly 5.5 dB in signal-to-noise ratio as compared to unmodulated noise. On the average nearly the same gain is obtained for competing discourse as a masker. However, for hearing-impaired listeners the gain with optimum modulation of the masker is only 1.2 dB, and for speech as a masker the speech-reception threshold is even raised by 3.2 dB. As a result the difference between the two groups of listeners increases from 2.7 dB for unmodulated noise to 10.5 dB for competing speech with the same character as the signal. For the hearing-impaired listeners a significant correlation (r = 0.79) is found between the SRT in unmodulated noise and the SRT with competing speech. No such correlation is found for the normal-hearing listeners. Nearly independent of the modulation frequency, for both groups of listeners a slightly higher SRT is found when a part of the masker spectrum is modulated in anti-phase.

The data for the three listeners with more severe hearing loss are not shown in Fig. 1, but they are essentially similar to the data of the other hearing-impaired listeners.

**DISCUSSION**

The data obtained here confirm the earlier results by Carhart and Tillman (1970) and by Duquesnoy (1983a) showing that hearing loss for speech is much larger when competing speech is used as a masker than for a steady-state noise masker. They also make it clear why the original high estimate (113) for the average growth of SHL in noise as a function of SHL in quiet (Plomp, 1978) is still useful as an indication for the handicap in everyday listening situations. When we try to measure the social handicap of an individual hearing-impaired listener, fluctuating noise seems more closely related to everyday listening condition than steady-state noise. If the high correlation between the results for these two maskers found for the hearing-impaired group in this study proofs to be reliable, results obtained with steady-state noise may be used to predict the social handicap.

In unmodulated noise the lowest SRT is found for modulation frequencies between 16 and 32 Hz in both groups of subjects. For lower modulation frequencies complete words may be masked with a detrimental effect on sentence intelligibility. For higher modulation frequencies the limited temporal resolution of the auditory system hampers the release from masking by the relatively silent intervals in the masker. The much smaller release from masking in hearing-impaired listeners compared to normal-hearing listeners must be caused by a reduced temporal resolution of the auditory system. Any substantial influence of the absolute threshold, by which the gaps in the noise would be less deep for the hearing-impaired listeners than for the normal-hearing listeners, can be excluded on the basis of the individual audiograms. For the hearing-impaired listeners and even for some of the normal-hearing listeners, speech appears to be a better masker for the sentences than noise. Duquesnoy (1983a) obtained equal thresholds for hearing-impaired listeners when masking sentences with noise or speech, but in contrast with our study, in his investigations the signal and the competing speech were from different speakers.

**REFERENCES**


IS HIGH-FREQUENCY HEARING NECESSARY FOR NORMAL LOUDESNESS GROWTH AT LOW FREQUENCIES?

Rhona P. Hellman and Carol H. Meiselman
Auditory Perception Laboratory
Northeastern University, Boston, MA 02115, USA

INTRODUCTION

In normal hearing, the loudness of a tone increases with sound pressure level up to at least 105 dB despite the simultaneous presentation of a masking noise adjacent to the tone on the high-frequency side (Hellman, 1974, 1978; Scharf, 1964; Zwicker 1958). To achieve these results, the sound pressure levels of the noise and tone are varied together so that the signal-to-noise ratio remains constant. Such experiments assume that high-frequency information available for the tone in the absence of noise is effectively contained by the noise.

Does the noise in fact limit the high-frequency spread of excitation evoked in the ear by the tone? To help answer this question, we measured loudness growth at low frequencies in cochlear-impaired ears with sloping high-frequency losses. Under these circumstances, the addition of noise is unnecessary. Any alterations in loudness that result can clearly be ascribed to the absence of high-frequency hearing.

Our goal was to determine whether loudness functions measured in cochlear-impaired ears with sloping high-frequency losses were analogous to those measured in normal ears partially masked by noise.

METHOD

Nine people, 21 to 69 years old, all with cochlear impairment, served as listeners in the experiments. Listening was monaural through a TDH-49 earphone mounted in an MX-41AR cushion. The criteria for selection of listeners was based on (a) the lack of an air-bone gap at frequencies in the region of hearing impairment; (b) evidence of a bilateral symmetrical sloping high-frequency hearing impairment; (c) thresholds that initially increase with frequency by 30 dB/octave or more beyond the normal-hearing range; (d) otological examination and history; (e) a diagnosis of cochlear impairment by conventional audiological tests. Thresholds for those listeners who met these criteria were evaluated in greater detail by the method of limits to determine the appropriate frequencies to be used for the loudness measurements.

A TDH-49 earphone supplemented by either a Koss or Yamaha-1 earphone were used for threshold measurements at very high frequencies.

Figure 1 shows the measured thresholds as a function of frequency for each of the nine listeners. The sound pressure levels (SPLs) in Fig. 1 are values corrected in accordance with the recommendations by Fastl and Zwicker (1983). The corrections are necessary to obtain more exact SPLs at the entrance to the ear canal. On the average, the corrected thresholds increased by 43 dB/octave beyond the chosen cutoff frequency.

The stimuli consisted of pure tones generated by a Hewlett Packard audio oscillator. Within the normal-hearing region two test frequencies, a minimum of one octave apart, were studied. The listeners, seated in a sound-proofed booth, judged loudness by absolute magnitude estimation (AMOE), and by absolute magnitude production (AMP) [see Hellman and Zwierlicki, 1965, 1966]. The loudness adjustments for AMP were made with a move potentiometer that covered a 70 dB range. To illustrate the concept of an open-ended numerical scale, AME of line length preceded AMP of loudness. Following the

Fig. 1. Threshold of audibility for nine listeners with sloping high-frequency cochlear losses.

threshold evaluations and the line length estimations, the loudness of each test frequency was judged in a separate run. Half of the listeners judged the loudness of the lower test frequency in a first run, and half judged the loudness of the cutoff frequency in a first run. The range of lower test frequencies was from 100 to 1000 Hz, and the range of cutoff frequencies was from 2000 to 3200 Hz. For reasons given elsewhere (Hellman and Zwierlicki, 1965), AME always preceded AMP. Loudness was measured at 4 dB sensation level (SL) and above, but only judgments at 30 dB SL and above where loudness is a simple power function of sound pressure were used for the computation of the individual loudness exponents.

RESULTS AND DISCUSSION

Figure 2 shows individual loudness functions measured at the lower test frequency. The selected frequency was determined from the threshold functions in Fig. 1. Two loudness functions are drawn for each listener, one for AME (solid line) and a second for AMP (intermittent line). The lines were determined by a least-squares fit to the data. Also shown are the individual geometric means of the final two loudness estimations (filled and unfilled circles). The SPLs differ from listener to listener because of individual threshold differences that were taken into account in

Fig. 2. Individual loudness functions measured at the lower test frequency in the normal-hearing region. Numbers in parentheses are the loudness exponents determined by the combination of AME and AMP. Arrows indicate the reference SPL of 80 dB for each pair of functions.
All the data in Fig. 2 are well described by power functions. As typically found, the curves resulting from AMP are steeper than the ones resulting from AME (Stevens and Greenbaum, 1966). Individual exponents (numbers in parentheses) are based on the geometric mean (GM) of the exponents for AMP and AME. The individual exponents range from 0.20 to 0.95 re sound pressure, but 5 of the 9 values lie within ±1 SD of the group GM of 0.61. Moreover, although the exponent for each pair of loudness functions was determined for a different test frequency, the group exponent corresponds closely to 0.60 which is the international standard for loudness (ISO R 131-1959).

Figure 3 shows individual loudness functions measured at the cutoff frequency, beyond which thresholds markedly increased with frequency. The plot is analogous to that in Fig. 2, except that the filled and unfilled circles are the results of AMP. Each point is the decibel average of the final two judgments. Both the range and average exponent values in the steep slope of the threshold functions at frequencies just above the cutoff frequency, the NMB functions at the lower test frequency (solid lines) do not noticeably differ from the ones at the cutoff frequency (interrupted lines). Both the dynamic range measured for loudness and the tone's rate of loudness growth are unaffected by the lack of high-frequency hearing.

The results indicate that the magnitude of loudness does not depend on the high-frequency spread of excitation of the tone. Much evidence in normal hearing (Hellman, 1974, 1978; Scharf, 1964; Zwicker, 1958) and in cochlear impairment (Hellman, 1983; Moore, Glaser, Hess, and Birchall, 1985) has accumulated to suggest this conjecture. Nonetheless, the results in impaired ears are not completely analogous to those in normal hearing partially masked by noise. Whereas in normal hearing the addition of noise does not alter the growth rate of loudness, noise typically decreases the slope of the loudness function (Hellman, 1978). No such slope decrease was observed either in the preliminary study or in the study of Moore et al. (1965) in impaired ears under "ramped conditions." Further experiments are needed to understand these results. They provide additional evidence that the effects of a masking noise in normal hearing are not equivalent to the effects of cochlear impairment.

ACKNOWLEDGEMENTS AND NOTES

1. The large regression angle between the curves of AME and AMP for listener WR is not unusual for psychophysical scaling tasks (S.S. Stevens, 1975).

This research was supported by the Rehabilitation Research and Development Service of the VA.

REFERENCES


PSYCHOACOUSTIC PERFORMANCE IN SUBJECTS WITH DISCRETE NOISE INDUCED HEARING LOSS

F. Lindgren

Dept. of Occupational Audiology, Sahlgren Hospital, S-413 45 Gothenburg, Sweden

INTRODUCTION

Pure tone audiometry is today a very common hearing examination. Unfortunately, this examination will only provide limited information of other dimensions of hearing than those related to the threshold. It has been shown that even a small pure tone hearing loss can correspond to a substantial loss of cochlear sensory cells (Hedberg, 1968). Consequently, it is of interest to develop tests that with accuracy and precision will confirm early or progressive noise induced hearing losses (NIHL).

Discrete NIHL has been shown to affect speech discrimination in noisy backgrounds (Antiasson, 1973). Thus, speech discrimination may be dependent upon time- and frequency resolution of the hearing sense. These factors are probably more influenced by noise exposure than revealed by the pure tone threshold. Earlier studies of relationships between psychoacoustic tuning curves (PTC) and speech discrimination report a stronger correlation between PTC and speech discrimination than between PTC and pure tone thresholds (Boding, 1979; Festen & Plomp, 1983).

Frequency discrimination, when measured as difference limens for frequency (DLF) has been shown to correlate with speech discrimination in individuals with sensorineural hearing loss (Risberg & Agefors, 1978; Gengel, 1973). It has also been reported that DLF as well as PTC were affected at test frequencies showing normal pure tone thresholds in subjects with a high frequency hearing loss (Turner & Nelson, 1982; Leshowitz et al., 1976). Substantial influence on PTC in individuals with pronounced sensorineural hearing losses are reported by several authors (Carney & Nelson, 1983).

Another parameter of interest is the temporal integration (TI). TI is usually measured as the difference in dB between two threshold determinations; one performed with normal length of the pure tone stimulus, and the other with reduced length of the tone stimulus. Chung (1982) reports that TI measured with Békésy-technique and 20 msec duration of the stimulus, is connected to the configuration of the audiogram in subjects with different degrees of NIHL.

While earlier investigations mainly concerned psychoacoustic performances in subjects with pronounced hearing losses, the aim of the present study was to compare psychoacoustic performances in normal hearing subjects and in subjects suffering from a discrete NIHL.

MATERIAL AND METHOD

Twenty-seven subjects, age 17 – 21, with normal pure tone thresholds (< 10 dB HL) at all test frequencies when measured with audiometer Interacoustics AC-4 in the frequency range 0.25 – 8 kHz constituted the control group (NG-group) while 11 subjects, age 17 – 22, with a maximum hearing loss of 20 or 50 dB HL at 4 or 6 kHz respectively, were selected to the experimental group (NIHL-group). All tests were carried out in a sound-proof booth.

All subjects were submitted to:
- Pure-tone sweep frequency Békésy audiogram performed on the selected test ear.
- Frequency discrimination (DLF) was assessed by a 2-alternative forced-choice decision paradigm. Each subject’s task was to indicate which tone of a temporally sequential pair was “higher”. Indicator lights informed the listener of the correct response at the end of each trial. DLF’s were measured at 1, 2 and 4 kHz and test tones were presented at 20 dB sensation level (SL). Seven frequency differences were tested twice during each experiment, providing seven data points on the subject’s psychometric function, each point based on 40 decisions. The value of the seven frequency differences were chosen in order to obtain a threshold corresponding to 75% on the psychometric function.
- Temporal integration (TI) was obtained when subjects tracked their thresholds twice by Békésy procedure in the entire frequency region 1 – 8 kHz. The procedure involved comparison of threshold levels for a 250 msec and a 20 msec tone pulse. Rise and decay time was 7.5 msec. Repetition rate was in both cases 2 pulses per second.
- Acoustic reflex threshold (ART) measurements were carried out with a Madsen ZO-72 bridge, using a 220 Hz probe tone. The ART was determined in dB HL of the contralateral stimulus tones as the first level where a significant shift in impedance was found. ART’s were in each subject determined at 0.5, 1, 2 and 4 kHz.
- Psychoacoustic tuning curves (PTC) were recorded with center frequencies 1, 2 and 4 kHz. The pulsed test-tone as well as the continuous pure tone masker were mixed and presented to the subject over a TDH-39 ear-phone. The test tone was in all cases presented at 10 dB SL. In each trial, the masked threshold was determined by Békésy-tracking of the masker level. The subject was asked to keep the test-tone just audible while ignoring the masker. The masker was presented at three frequencies below as well as three frequencies above the test tone frequency. The masker frequencies had the following proportions to the center frequency: 0.43, 0.78, 0.92, 1.08, 1.23 and 1.48. The PTC-curves were normalized (i.e., individual TTL at test tone and masker frequency was subtracted from all masked levels) giving curves with coinciding levels of test tones for both groups.

RESULTS

Analysis of differences between the two groups, showed that there were statistically significant differences in hearing thresholds at 0.25, 3, 4, 6 and 8 kHz (p < 0.05) (Table 1). Comparison of differences as in TI values were not statistically different between groups at any specific test frequency (Table 2). However, DLFs were significantly better (p < 0.05) in the NIHL-group at 4 kHz (Table 3) while ARTs were higher at all test frequencies in the NIHL-group (p < 0.05) (Table 4). Comparison of PTCs, revealed an elevated masker level in the NIHL-group at the three test frequencies. This elevation was verified as significant at masker frequencies 0.43, 0.78 and 1.0 kHz (test tone 1 kHz) (Fig. 1) and at masker frequencies 1.72, 3.12 and 3.68 kHz (test tone 4 kHz) (Fig. 2).

DISCUSSION

The results from this study support the hypothesis that psychoacoustic performance is influenced by moderate NIHL. The unexpected
difference in ARTs at frequencies where HTLs were equal, can have a significant importance in designing tests for early detection of sensorineural hearing loss. The decreased frequency resolution demonstrated by PTC at 1 kHz could also be an indicator of damage at this frequency, in spite of the fact that this is not yet seen in the audiogram. Further analyses with larger experimental groups might confirm this. The fact that DLFs were affected only at 4 kHz suggests that DLF is in a higher degree related to the HTL than to other factors, and might be rather insensitive at frequencies located at some distance from the injury.

Further analysis of the material, including regression and correlation analysis will be performed in order to study the relative importance of the tested psychoacoustic parameters on hearing thresholds at affected and unaffected test frequencies.

TABLES AND FIGURES

---

**Figure 1.** Normalized psychoacoustic tuning curves at 1 kHz in NH-group (broken line) and in NIHL-group (unbroken line). Vertical bars indicate one standard error of the mean.

---

**Figure 2.** Normalized psychoacoustic tuning curves at 4 kHz in NH-group (broken line) and in NIHL-group (unbroken line). Vertical bars indicate one standard error of the mean.

---

**Table 1.** Pure tone thresholds in dB HL, mean and standard deviation in 27 normal hearing (NH) subjects and in 11 subjects with slight NIHL.

<table>
<thead>
<tr>
<th>Frequency in kHz</th>
<th>0.25</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>1.7</td>
<td>-0.1</td>
<td>-0.1</td>
<td>0.4</td>
<td>-1.0</td>
<td>1.1</td>
<td>0.9</td>
<td>-1.2</td>
</tr>
<tr>
<td></td>
<td>5.4</td>
<td>4.3</td>
<td>4.7</td>
<td>4.6</td>
<td>5.5</td>
<td>4.6</td>
<td>6.0</td>
<td>5.7</td>
</tr>
<tr>
<td>NIHL</td>
<td>6.4</td>
<td>2.7</td>
<td>2.3</td>
<td>3.2</td>
<td>7.8</td>
<td>9.1</td>
<td>30.0</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>3.9</td>
<td>5.6</td>
<td>6.1</td>
<td>6.8</td>
<td>6.1</td>
<td>6.3</td>
<td>9.3</td>
<td>5.5</td>
</tr>
</tbody>
</table>

**Table 2.** Temporal integration (TI) in dB, mean and standard deviation in NH and NIHL group.

<table>
<thead>
<tr>
<th>Frequency in kHz</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>5.8</td>
<td>5.2</td>
<td>5.5</td>
<td>7.3</td>
<td>7.7</td>
<td>3.8</td>
<td>6.0</td>
<td>7.2</td>
</tr>
<tr>
<td></td>
<td>2.4</td>
<td>2.2</td>
<td>3.2</td>
<td>4.5</td>
<td>3.7</td>
<td>4.0</td>
<td>6.0</td>
<td>4.5</td>
</tr>
<tr>
<td>NIHL</td>
<td>7.1</td>
<td>7.0</td>
<td>6.3</td>
<td>5.6</td>
<td>7.6</td>
<td>2.1</td>
<td>3.5</td>
<td>4.5</td>
</tr>
<tr>
<td></td>
<td>3.7</td>
<td>3.5</td>
<td>2.2</td>
<td>3.4</td>
<td>3.5</td>
<td>5.9</td>
<td>4.9</td>
<td>4.5</td>
</tr>
</tbody>
</table>

**Table 3.** Frequency discrimination (DLF), mean and standard deviation in NH and NIHL group.

<table>
<thead>
<tr>
<th>Frequency in kHz</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH</td>
<td>90.2</td>
<td>90.0</td>
<td>90.2</td>
<td>93.3</td>
</tr>
<tr>
<td></td>
<td>7.5</td>
<td>6.7</td>
<td>5.8</td>
<td>12.7</td>
</tr>
<tr>
<td>NIHL</td>
<td>98.0</td>
<td>97.0</td>
<td>97.0</td>
<td>107.0</td>
</tr>
<tr>
<td></td>
<td>11.6</td>
<td>12.7</td>
<td>10.9</td>
<td>9.5</td>
</tr>
</tbody>
</table>

**Table 4.** Acoustic reflex threshold (ART), mean and standard deviation in NH and NIHL group.

---

REFERENCES

BASES FOR THE ANALYSIS OF PSYCHO-SOCIAL DISADVANTAGES DUE TO NOISE-INDUCED HEARING LOSS

Monique Lalonde, Raymond Héroux

Département de santé communautaire de l'Hôpital Saint-Luc, Programme régional-audiologue, 1001, rue Saint-Denis, Montréal, Canada, H2x 3H9; et GAUM, Université de Montréal, C.P. 6128, Montréal, Québec, Canada, H3C 3J7

INTRODUCTION

The World Health Organization has proposed an operational definition of the concepts of impairment, disability and handicap related to disease (1). This conceptual scheme was applied by Davis, among others, to hearing problems, in 1983 (2). It allows the consideration of hearing problems as a function of the affected individual and his environment.

The clear differentiation between the concepts cited above was particularly helpful in the field of deafness. Indeed, until recently, the disabilities and handicap concepts have been generally mistaken. Unfortunately, people have been referring too often to the American medico-legal definitions which regrupps all "handicaps" into one disability and describe the disability from a strictly administrative point of view as "the inability to remain at full wages" (3). Because of the ambiguity in these concepts, researchers have been mainly concerned with the disabilities caused by sensori-neural impairments; as a result, psycho-social disadvantages (handicaps) have been very little investigated. However, to reduce these disadvantages, we need to identify them, to know when they become manifest, and to be able to evaluate them.

The aim of the present study was to develop a questionnaire that permits an exploration of the mutual adaptations in the family with respect to the disabilities caused by noise-induced hearing loss (NIHL) and to evaluate the psycho-social disadvantages experienced in that context. The study of family interactions has been considered in the perspective of future rehabilitation programs that must take into account the causes of the handicap.

EXPERIMENTAL APPROACH

The prevention or reduction of the psycho-social disadvantages due to an occupational hearing loss undoubtedly depends on the use of appropriate behavioral responses by both the individual and his surroundings to inconveniences caused by the partial deafness. To acquire such behaviour, they must first become aware of the hearing problem and acknowledge it. A good understanding of the consequences of the NIHL also appears to be very important at this point. It is in this perspective that the questionnaire has been constructed. This questionnaire has been derived from a previous analysis of this particular subject (4).

In addition to one question on the inability to understand speech in the presence of background noise, the questionnaire includes four categories of questions concerning the following subjects:

1. The acknowledgment of the hearing problem (3 questions).
2. The means by which the affected individual coped with the impairments and the disabilities associated to the NIHL (5 questions) and to the stress caused by exposure to noise (3 questions).
3. The remarks, reproaches and/or expressions of annoyance from the family to the limits imposed by the hearing loss (6 questions).
4. The number of discussions related to NIHL (2 questions).

The experimental questionnaire was submitted to a population of electrical motor shop workers who had taken part to a hearing test. The survey was made according to the procedure recommended by Héroux et al in 1983 (5). In order to see whether the questionnaire was understood by the subjects and to improve the interpretation of the answers, interviews were undertaken in parallel with twelve workers suffering from considerable occupational hearing loss. At the end of interview, the investigator asked the workers to give on a scale their opinions about possible rehabilitation services that could be made available to them.

STUDY SAMPLE

The answers to the questionnaire were analysed for about a hundred workers, only men with a mean age of 50.2 years who had either normal hearing for their age or a hearing impairment that was characteristic of NIHL. The workers were asked to complete the questionnaire with their spouse or a close relation. This instruction was followed in about 90% of the answerers.

In order to evaluate what the experimental questionnaire was measuring, a factor analysis (principal components model with varimax rotation) was used after binary coding of the nominal answers. The variables included in the computation were the respondent's age, hearing evaluation, number of children living at home and all other questionnaire items except for three questions for which the distributions were clearly asymmetrical. The hearing evaluation was also introduced using the average threshold for 1.2, 3, and 4 kHz from the worst ear (7).

RESULTS

The correlation coefficient matrix shows that 18 of the 23 variables correlated significantly with the workers' perception of having a hearing problem. Only three variables made exceptions: respondent's age, number of children living at home, the fact of providing an explanation after asking for quiet at home, after work. Furthermore, it is noteworthy that except for these items, there was a very large number of statistically significant intercorrelations between the variables.

In order to visualize the complex structure of the interactions between the variables, a principal component analysis was performed. The resulting loadings matrix defined seven common factors among which the first three explain 76.4% of the variance (Table 1). The careful examination of their composition allows to give them the following labels:

Factor 1: Sensation of a hearing problem
Factor 2: Awareness of imposing constraints
Factor 3: Feeling of being inadequate.

The questions referring to the family reaction were distributed in the different factors. They can be divided into two groups: a) items 7, 10, 15, 16 are direct questions related to remarks and reproaches from the family; b) questions 6, 8, 9, 11 probe into the worker's perception with respect to the expression of annoyance from the part of the family.

The first factor in the analysis reveals that the subjective perception of a hearing problem is mostly determined by the disability to discriminate speech in background noise (Q12) and the necessity to resort on optical mechanisms (Q10, Q18). The family reaction also contributes to that perception but
to a lesser degree and it appears that it is more the question 4 (implicating the comments coming from a more enlarged surroundings) that is determinant at that level.

The contribution of the family reaction to the perception of the hearing was further examined by means of discriminant function analysis. The questions 7, 10, 15, 16 of the dependent variables and the audiometric formula selected, the independent variable. The result indicates that only remarks about the television and the voice being too loud are more sensitive for the identification of a hearing problem than the worker's own perception. However, the comments about a too loud voice have no specificity while the reproaches addressed to the worker are highly specific (Q15 = 93.7%; Q16 = 98.8%) but little sensitive (Q15 = 31.3%; Q16 = 38.3%).

Although the family influence does not seem to be determinant to the awareness of a hearing problem, it is different for the Factors 2 and 3. The family response to the disabilities related to occupational hearing loss and to the coping behaviours of the affected worker, strongly contribute to their impression of him/her imposing constraints and being inadequate.

Five questions relating to the family reactions present very significant loadings on Factor 2. Except question 7 about loud voice, the other questions concern expressions of annoyance that the affected person says he/she perceives in their family (intolerance, impatience, irritation...) and that they have provoked by asking to repeat (Q6), by talking loud (Q8 and Q9) and by setting the radio and television volume too loud (Q11).

The feeling of being inadequate associated with Factor 3 seems to depend mainly on family reproaches (e.g. question 15 and 16) and on the number of children living at home. In this factor are also included the expression of annoyance felt when being asked to repeat (Q6) and the misinterpretation of aggressiveness resulting from a loud voice (Q9).

DISCUSSION

The experimental questionnaire, by accounting for family reaction, permitted to study the contribution of close relations to the identification of a hearing problem and the actualization of the psycho-social disadvantages.

The factor analysis was not performed on many samples to confirm the stability of the configuration of the principal-components and the number of variables submitted was large with respect to the sample size. The results nevertheless obtained give some basic guidelines for the development of rehabilitation programs for people suffering from NHHL. They put forward the need to involve the family in such programs. Indeed, they suggest that, left to itself, the family tends to act as a catalyst of psycho-social disadvantages. Furthermore, the family does not seem to help significantly the worker to acknowledge his/her hearing problem. Actually, in a recent study based on interviews of noise-exposed workers (8), the family integration difficulties appeared to represent one of the disadvantages that the affected individual feels the most acutely.

The information reflected in the present study by means of interviews, has led to similar conclusions. On the one hand, the participant's comments reflect the fact that the hearing loss remains a taboo at home and, on the other hand, this hearing loss is the root of important constraints. The spouses do not explicitly look for mutual acceptable solutions to the consequences of the hearing loss. Furthermore, the workers have clearly indicated that they felt the need that their spouse understood their lived experience with hearing loss. Consequently, the family seems to play a very crucial role with respect to the adaptation to noise-induced hearing loss. This should be taken into account in any rehabilitation approach, by helping the family to minimize or prevent the psycho-social disadvantages of the hearing loss.

Table 1: Factor loadings obtained with the principal-component factor analysis on the responses to the questionnaire

<table>
<thead>
<tr>
<th>FACTOR 1</th>
<th>FACTOR 2</th>
<th>FACTOR 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 0.74*</td>
<td>Q8 0.77*</td>
<td>Q16 0.69*</td>
</tr>
<tr>
<td>Q12 0.62*</td>
<td>Q9 0.66*</td>
<td>Q15 0.63*</td>
</tr>
<tr>
<td>Q12 0.60*</td>
<td>Q7 0.53*</td>
<td>NF 0.51*</td>
</tr>
<tr>
<td>Q4 0.57*</td>
<td>Q6 0.46*</td>
<td>Q3 0.40*</td>
</tr>
<tr>
<td>Q5 0.56*</td>
<td>Q5 0.38*</td>
<td>Q6 0.32*</td>
</tr>
<tr>
<td>Q9 0.30*</td>
<td>Q11 0.37*</td>
<td>Q9 0.27</td>
</tr>
<tr>
<td>Q19 0.33*</td>
<td>Q12 0.32*</td>
<td>Q12 0.27</td>
</tr>
<tr>
<td>Q3 0.29</td>
<td>Q12 0.26</td>
<td>PH 0.26</td>
</tr>
<tr>
<td>Q7 0.29</td>
<td>Q3 0.25</td>
<td>S012 0.26</td>
</tr>
<tr>
<td>Q13 0.27</td>
<td>(p&lt;0.05)</td>
<td>*p&lt;0.01</td>
</tr>
<tr>
<td>Q15 0.26</td>
<td>PH 0.23</td>
<td></td>
</tr>
</tbody>
</table>

REFERENCES


QUALITATIVE ANALYSIS OF PERCEIVED HEARING
DISABILITIES AND HANDICAPS AMONG NOISE-EXPOSED
WORKERS.

L. Girard, R. Héto, L. Getty.

Institut Raymond-Dewar, 3600 Berri, Montréal and
Groupe d'acoustique, Université de Montréal,
2375 Côte Ste-Catherine, Montréal.

Noise-induced hearing loss (N.I.H.L.) is the most
prevalent occupational disease. Few affected
workers consult, and when they do, it is gener-
ally for a compensation claim. They rarely have
access to rehabilitation services.

In addition, noise-induced hearing-impaired
workers may not benefit from classical rehabilia-
tion programs because a) their impairment appears
very progressively and is frequently associated
with getting older (1), b) the auditory fatigue
causes daily fluctuations in hearing (2), c) the
people at work living the same difficulties seem
to a person to perceive his own disabilities as normal
(3), d) there are negative opinions in the society
toward hearing-impaired persons (4), and e) most
of these persons can not use hearing amplification
(1), because of high noise exposure.

Factors that could motivate these people to
participate in a rehabilitation program as well
as the components of such program are still to be
known. There is a need to identify the critical
conditions for an effective rehabilitation of this
specific group of hearing-impaired persons.

Thus, a study was conducted with the intention
of answering the following questions: 1) what are
the disabilities and disadvantages perceived
by the most undesirable by affected persons? and 2)
to what variables are they associated?

METHOD

Subjects
The sample comprised 26 men aged between 30
and 64 (mean: 48.1 ± 8.1). They were all employees
of a bottling firm, working in the maintenance de-
partment for the most of them (65%), they repre-
sented 85% of the workers in this department. Thus,
this sample is representative of a working group.
They had been exposed to noise at work from 7 to
40 years (mean: 21 ± 13.1).

Procedure
Subjects had their hearing tested and answered
a short questionnaire. An interview was then sched-
uled. This interview was done according to a pro-
tocol which comprised the following: a) identifi-
cation of every situation where a hearing disabili-
ity is being felt, b) evaluation of the negative
impact involved in each of them, c) identification
of the coping strategies, the manifest consequen-
ces and the possible social disadvantages, and d)
identification of the most disturbing difficulty.

The hearing reports were analysed according to
Héto et al. (5). The content of every recorded
interview was set into a written form and submitted
to a content analysis (6).

RESULTS

Two different categories of disabilities and
disadvantages (7) had been identified: those common
to a majority of interviewed persons and those ex-
pressed by just a few of them, and related to their
personnal characteristics. Hence, 19 persons (73%
of the group) reported communication problems in
noisy environments. Most of them also reported
making efforts to maintain satisfying conversation.
Because of these problems, some felt they were dis-
turbing to their near relations; this in turn af-
fected their self-esteem. On the other hand, only
one person suffered from not being able to listen
and enjoy classical music anymore. This is an ex-
ample of a hearing disability whose manifestations
depend on individual life style; the disadvantage
is nevertheless felt strongly. In short, general
and idiosyncratic hearing handicaps resulting from
N.I.H.L. had been reported; disabilities affecting
social relationships and disadvantages affecting
self-esteem were identified as the most undesira-
able handicaps.

The effects of different variables on the han-
dicaps resulting from N.I.H.L. were also analysed.
The results showed that age is important and con-
tributes to modify the perceived handicaps. In
fact, the communication problems in noisy environ-
ments and the consequences related to this disabi-
ility (discomfort, feeling of disturbing or of lock-
ing unusual, etc.) were reported by a majority of
persons, regardless of their age. However, the
persons whose age was 50 years and older attenuated
these consequences, saying they "were getting old"
or they "were not 20 years old anymore"; they per-
ceived their own behavior in these situations as
being similar to those of their age group. It is
agreed, in the general population, that older peo-
ple are frequently hearing-impaired and behave con-
sequently in noisy environments: it is even per-
ceived as normal. Thus, a person having a
N.I.H.L. who is approaching retirement can perceive
himself or herself as normal, as a member of the
older community. In this context, the disability
would not affect the self-esteem.

On the other hand, the younger persons (30
to 60 years old) affected by N.I.H.L. could hardly
attenuate the manifest consequences likewise. Some
compared themselves to other people at work, but
here the comparison group is much smaller and not
well known in the general population. Thus, they
feel more disadvantaged.

In addition, every interviewed person aged
between 30 and 60 years old reported anxiety toward
the fact that their hearing could worsen in the future.
It can be interpreted as if they were really concerned about the effects, or possible
effects, of noise on their hearing.

DISCUSSION AND CONCLUSION

The disabilities associated to N.I.H.L. are
relatively well known and documented. This study
lead to the conclusion that the most undesirable
disabilities, according to interviewed workers,
concerns social relationships. It is not surprising
since hearing is involved in oral communication;
however, it implies that a rehabilitation program
for these specific group of persons should, as much
as possible, in priority, aim at helping them im-
proving their communication skills in background
noise. It also implies that N.I.H.L. has probably
some effects on the near relations of the affected
workers, which means that their family would also
benefit from taking part to the rehabilitation pro-
cess.
A rehabilitation program could not ignore the disabilities that affect only a small number of individuals because of their lifestyle. Their motivation for improving their condition may be very strong in view of the particular disadvantage they feel.

The social disadvantages associated with N.I.H.L. are not as well known as the disabilities. This study indicated that those related to self-esteem were the most undesirable, according to the analysis of the interviews. It also showed that the younger workers (30 to 40 years old) were probably more disadvantaged and felt more anxious about their hearing. These observations imply that a rehabilitation program would probably, in a first step, give more benefit to young exposed workers than to their older colleagues who are likely to associate themselves with the older community. Stronger motivation and higher perceived benefits would be felt by the younger workers. These observations also indicate that a rehabilitation program should help in restoring self-esteem.

Rehabilitation should likewise be offered to the older affected workers who, even if they have found ways to cope with their disabilities, sometimes live serious difficulties. For such an intervention to be effective, factors affecting motivation to participate in rehabilitation should be further identified in order to make possible the initiation of such programs.

The present study was exploratory and qualitative in nature. The major findings should be confirmed on a large sample of workers by means of a questionnaire involving a limited number of items. It nevertheless offers valuable clues in identifying the critical conditions for an effective rehabilitation of the people suffering from occupational hearing loss.

REFERENCES


7. U.M.S. Classifications internationale des handicaps: déficience, incapacités et désavantages. INSERM /s.d./.
COMPUTERIZED AUDIOMETRY STATION

E. Colomb, J.C. Ventura, A. Pegna
Institut de Physiologie, R. du Bugnon 7, 1006 Lausanne, Switzerland

Introduction

With the fascinating development of research in the fields of cochlear physiology and VIIIth nerve neurphysiology, the traditional audiological tests have become obviously insufficient. The audiologists need a much more versatile audiometry station to set up new tests.

The standardization of the IEEE-488 bus in laboratory equipments allows an easy device interfacing for the assembling of a computerized audiometry station.

To allow the implementation of any psychoacoustical test, we hypothesized that two types of stimuli are sufficient: pure tones and width controllable noise bands. Having these available, we must be able to control their timing, amplitude and frequency location.

In order to achieve these functions the following apparatus are gathered:

1. A sine generator.
2. A white/pink noise generator.
3. A programmable filter network.
4. Programmable attenuators for each signal.
5. A mixer cascaded to an attenuator that sets the overall stimuli dynamic.
6. A programmable reed-relay array.
8. A dialog unit that allows the coding of subjects' answers by using a yes/no (louder/softer) switch.
9. An AM-SC modulator. It synthesizes a pair of noise bands out of a baseband one, using the sine as carrier.
10. A HP85B desktop computer for the station control.
11. A floppy disk unit for storing software and acquired data.

These devices are assembled according to fig. 1 functional diagram.

At present, the implanted software performs measures of the auditory threshold (AT), masking patterns and width of critical bands (CB).

Auditory threshold

Two methods are used for this test. One measures the AT at discrete frequencies separated by 1/3 of an octave, ranging from 125Hz to 8kHz. The patient receives a continuous pure tone at each measured frequency. He is instructed to indicate if he hears the tone or not, by pressing the yes/no switch. He thereby increases the level by 10dB steps until detection occurs, after which it is decreased by 5dB steps. Once perception is lost the level reincreases by 1dB steps. The latest perception level is then stored as being the AT at its frequency (Fig.2). In the second method (Bekesy) the AT is tracked in a virtually continuous frequency sweep. The sine generator is set at 100Hz and frequency is continually increased from here by 0.005 octave steps at a rate of 3.6 steps/second. Using the dialogue unit, the patient's answers entail an increase or a decrease in the tone level of 2dB/step. This means that the tone cannot be maintained at a fixed level. In this way the tone alternates between perception and non-perception points. An example of an obtained curve is given at fig. 3-A. By performing an averaging over the recorded samples the AT is obtained, as shown in fig. 3-B.

Although the represented curves are given in dB SPL, they can be transformed to dB SL diagrams by simply subtracting the samples of a normalized threshold to the measured ones (a routine for this purpose is implanted).

Another routine yields a graph representing the average of the stored ones. At each run this routine loads two files, one with the average of a number of curves and the other with a new curve. A new average is then recalculated and stored.

Masking pattern

The tracking of masking patterns proceeds as for the Bekesy AT with the addition of a masking tone at a fixed chosen frequency. In order to avoid beats in the masking frequency's vicinity, the used tone is narrow-band noise instead of a pure one. The minimal allowed width of the bandpass filter (7% of center frequency) is used for masking frequencies over 1kHz whilst the bandwidth for lower frequencies is of 10% of center frequency. This band enlargement at low frequencies is made to avoid an audible pulsing of the masking tone. Fig. 4-A shows a masking pattern diagram superposed to the corresponding auditory threshold. Fig.4-B represents these diagrams' difference.
Experiments reported on reference [2] mention a phenomenon called "Pronounced Asymmetry of Masking" (PAM). PAM is characterized by an unusual enlargement of the masked region towards the higher frequencies. Though such an asymmetry exists for normal hearing subjects, it does not take such important proportions. No correlation however has been sought between PAM and speech intelligibility in patients. This audiometry station should aid in a more systematic analysis of this phenomenon.

Critical bands

According to the CB model, loudness of a complex auditory stimulus is perceived as the combination of partial loudnesses measured in several bandpass zones (these being called critical bands) resulting from a split operated by the ear over the whole spectrum. Each of these partial loudnesses is measured by integrating the spectral portion within the corresponding CB. In order to measure the width of a CB in a given frequency area, the three following properties are considered:

1. The auditory threshold of a pure tone masked by a noise band remains constant as long as both stimuli are within the same CB.
2. Critical bandwidth is independent of intensity.
3. The ear forms a CB, placing it so as it will contain a maximum amount of energy.

The first property suggests tracking the AT of a pure tone as a function of the gap between it and the noise band. Yet, the corresponding CB's position won't be steady during the test, due to the third property. Therefore, the method of the pure tone in a frequency gap [1] is used: the AT of a pure tone of fixed frequency is measured with the same "perceived/not-perceived" technique. This pure tone at $f_p$ is placed between two noise bands whose gap $\Delta f$ increases gradually (fig. 5). As the experiment begins, the pure tone is placed in the same CB as one of the noise bands (fig. 5-A). At some point the pure tone and each noise band should fall into separate CBs (fig. 5-C). At this moment the AT should start decreasing. The level of the tone, as a function of $\Delta f$, should therefore show a flat response curve followed by a sudden decreasing curve (fig. 5-A). The $\Delta f$ corresponding to the curve's breakpoint is the width of the CB. A routine exists for the averaging of the curve (fig. 7-B).

The pair of noise bands is synthesized from a single one at low frequency using the AM-SC modulator. The sine generator provides for both, the pure tone and the modulation carrier. In the baseband the lower cutoff frequency $f_c$ equals $\Delta f/2$. After frequency shift, the higher frequency noise band is supposed to be $1/3$ octave wide. This imposes the baseband upper cutoff frequency:

$$f_c = 2^{-2/3} \Delta f + (2^{1/3} - 1) f_p$$

As $\Delta f$ increases, the programmable filter changes the modulating signal cutoff frequencies. This brought an inconvenience to the experiment as the filter switching is audible. By placing a low-pass filter between the programmable one and the modulator most of the parasite switching "clicks" are suppressed. Unfortunately, for $f_p > 1$ kHz, the modulating signal band is broader and the clicks become audible. The disturbance it causes on some subjects requires a measuring procedure at discrete values of $\Delta f$ instead of a continuous broadening of the frequency gap.

Conclusion

At present, tests are being carried out on normal hearing subjects and on hearing impaired patients. The results on the former will yield a normalisation of the procedures to which the latter's results will be referred.

The accumulated data is planned to be transferred to a more powerful computer (PDP-11) for overall statistical analysis.

On the other hand, new experiments are being developed (psychophysical tuning curves).

Since the tested phenomena are of consequence to speech intelligibility, this research is necessary to the development of more performant and efficient hearing aids.

Acknowledgements

This achievement is part of a research project carried out in the laboratory of electromagnetism and acoustics (LEMA) together with the physiology institute of the Lausanne University and Swiss industries. It is financed by the Swiss commission for the encouragement of scientific research (CERS).

References

PROTECTION OF EARS DURING SELF-UTTERANCE
- A CONSIDERATION ON BONE-CONDUCTED VOICE -

KONO, Shunichi, TANAKA, Hiroshi and SONE, Toshio

Res. Inst. of Electr. Com., Tohoku University
2-1-1 Katahira, Sendai, 980 Japan

1. INTRODUCTION

It is reported by several authors that some protection effect against intense sound is observed in hearing during utterance. But its mechanism is still unclear. Middle-ear muscle contraction is considered to be one possible mechanism for protecting an inner ear against an intense sound. Since the energy of bone-conducted sound is great during utterance, the mechanism such as middle-ear muscle contraction may work for protecting ears against it. From this point of view, we studied the effect of bone-conducted sounds on TTS caused by external noise.

2. METHOD OF EXPERIMENT

In order to observe a TTS at 4 kHz, a band-limited pink noise (425 - 3400 Hz) including a critical band (2400 - 3400 Hz) for 4 kHz was used as a noise stimulus for the test. This noise stimulus was given to a subject and its level was 100 dB SPL at the subject's ears.

The hearing threshold of the left ear of a subject at 4 kHz was determined before and after exposure to noise. Hereafter, we will call the former threshold the "pre-exposure threshold" and the latter the "post-exposure threshold". The pre-exposure threshold was measured three times prior to each experiment by using an intermittent signal. Just after the exposure, the subject put on headphones and the post-exposure threshold was measured over a three minute period. Bekesy-type audiometer was used for the threshold measurement, and the TTS curve smoothed through a low-pass filter was observed. The post-exposure threshold was measured until 180 sec after noise exposure, and the average TTS for 2 minutes from 60 sec to 180 sec was determined. The average TTS's are indicated by TTS's and they were compared with each other for different conditions of experiment.

3. ALLEVIATION OF TTS BY UTTERANCE

Fig.1 shows the examples of TTS when the subjects uttered /a:/ sound at the level of 90 dBSPL at his ear during exposure to 100 dB noise and when they did not utter. These curves show the average TTS for five trials, and the ranges denote the standard deviations.

The subjects were four young men, of which two persons (Sub.1 and 2) show the alleviation of TTS by utterance. Those values were 1.9 dB and 3.7 dB respectively. On the other hand, the alleviation of TTS by utterance was 1.2 dB and zero in Sub.3 and 4. They are considered to be people hard to suffer from TTS, primarily.

4. ALLEVIATION OF TTS BY BONE-CONDUCTED SOUND

As to the hearing protection during utterance, Bekesy stated the effect of virtue vibration of stapes caused by a vertical vibration of a skull during utterance, and Karlovich assumed that the bone-conducted vibration transmitted to a skull caused middle-ear muscle contraction. However, we cannot find a study which ascertained directly whether a bone-conducted sound causes a hearing protection or not.

4.1 Alleviation of TTS by a synthesized bone-conducted vowel

A continuous vowel /a:/ synthesized after the real voices was given to the subjects through a bone-conduction vibrator attached to their forehead at a force of 5.4 [N]. Vibration acceleration level of a bone-conducted sound measured at a mastoid was the same as those observed during utterance. Fig.2 shows the spectrum of a bone-conducted sound expressed by vibration acceleration level. The results of the experiment were plotted on Fig.1.

4.2 Alleviation of TTS by bone-conducted noise

A band limited white noise, which has the component from 90 to 1800 Hz, was supplied as a bone-conducted sound. This frequency band is similar to that of the vowel /a:/ used in the previous experiment. The vibration given to the forehead was adjusted to the same level as that of the voice. Table 1 shows TTS's and their standard deviations obtained here and in the previous experiment.

The average alleviation of TTS for Sub. 1 and 2 was 2.8 dB for the utterance, 3.3 dB for the bone-conducted /a:/ sound, and 3.7 dB for the bone-
conducted noise. It is shown, therefore, that the alleviation of TTS is caused by bone-conducted sounds as well as the utterance.

Table 1. TTS'm's and the standard deviations around them for three conditions of exposure to noise: ( ) : standard deviation, ** : significant difference beyond 0.01 point

<table>
<thead>
<tr>
<th>SUB.</th>
<th>FATIGUE-NOISE ONLY</th>
<th>WITH /s/ WITH BONE CONDUCTED /s/</th>
<th>WITH BONE CONDUCTED NOISE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.1 (0.6)</td>
<td>6.2 (0.6)</td>
<td>4.8 (1.6)</td>
</tr>
<tr>
<td>2</td>
<td>11.8 (0.8)</td>
<td>8.1 (1.9)</td>
<td>8.3 (1.0)</td>
</tr>
<tr>
<td>3</td>
<td>5.6 (1.4)</td>
<td>5.4 (0.6)</td>
<td>6.1 (0.7)</td>
</tr>
</tbody>
</table>
| 4    | 4.5 (1.2)           | 4.5 (1.6)                        | 5.2 (1.0)                 | **

4.3 Change in the alleviation of TTS due to the difference in a frequency band of a bone-conducted noise

In order to investigate the effect of the difference in frequency of a bone-conducted sound, the frequency band of noise (90 - 1800 Hz) was divided into two. One of them has the frequency range from 90 to 400 Hz and the other from 900 to 1800 Hz. The level of vibration given to the forehead of subject was 3 dB lower than the original noise for both sub-band noises. The results of the experiment is shown in Table 2 where we can see that TTS'm's are alleviated to the same degree for both noises.

Table 2. Relation between alleviation of TTS and the frequency band of bone-conducted noise: ( ) : standard deviation, ** : significant difference beyond 0.01 point *

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>90 - 900Hz</th>
<th>900 - 1800Hz</th>
<th>90 - 1800Hz</th>
</tr>
</thead>
</table>
| 1       | 1.5 (0.9)  | 1.3 (1.0)   | 3.2 (1.2)   | **
| 2       | 0.4 (0.8)  | 1.8 (1.1)   | 4.2 (0.7)   |

According to the results shown in Table 2, there is a slight difference between TTS'm's for noises with both frequency bands in Sub. 1, while only the noise of 900 - 1800 Hz shows a significant effect on TTS in Sub. 2. Anyway, the broad band noise (90 - 1800 Hz) gives the greatest alleviation in TTS.

4.4 Change in the alleviation of TTS due to the difference in level of vibration

Comparing the results for three kinds of noise from Table 2 with each other, we cannot conclude as to which is effective for the increase in alleviation of TTS, the change in a frequency range or the increase in a level of bone-conducted vibration. Therefore, we obtained the alleviation of TTS for three levels of bone-conducted noise, the standard level (0 dB), -3 dB and -6 dB.

From the above results, we can see that the alleviation of TTS is greater for higher level of bone-conducted noise in Sub.1, while there is little difference in TTS between -3 dB and -6 dB of bone-conducted noise in Sub.2.

Table 3. Relation between alleviation of TTS and the level of bone-conducted noise: ( ) : standard deviation

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>0 dB</th>
<th>-3 dB</th>
<th>-6 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.2  (1.2)</td>
<td>2.0  (0.4)</td>
<td>1.1  (0.3)</td>
</tr>
<tr>
<td>2</td>
<td>4.2  (0.7)</td>
<td>2.4  (0.9)</td>
<td>2.7  (1.0)</td>
</tr>
</tbody>
</table>

5. RELATION BETWEEN THE FREQUENCY RANGE OF FATIGUE NOISE AND THE HEARING PROTECTION CAUSED BY BONE-CONDUCTED SOUND

The fatigue stimulus used hitherto is a band-limited pink noise which has a frequency band from 425 to 3400 Hz, of which the frequency band influential in TTS at 4 kHz is one from 2400 to 3400 Hz. In accordance to the critical band theory for TTS, dividing the above frequency band at 2400 Hz into two, we gave those sub-band noises to the subjects. Bone-conducted sound was the same band-limited noise as before. The sound level of fatigue noise was -0.8 dB for a noise from 425 to 2400 Hz and -7.8 dB for that from 2400 to 3400 Hz as compared with the level of original noise, as the spectral level of noise was fixed for three noises.

Comparing the TTS caused by an original fatigue noise with those caused by two sub-band noises, it was shown that TTS at 4 kHz was surely caused by the component within a critical band (2400 - 3400 Hz). In spite of this fact, the alleviation of TTS caused by these noises shows the values in Table 4.

Table 4. Alleviation of TTS for three kinds of fatigue noise: ( ) : standard deviation

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>425 - 3400Hz</th>
<th>425 - 2400Hz</th>
<th>2400 - 3400Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.4 (0.6)</td>
<td>-0.1 (1.0)</td>
<td>1.6 (1.0)</td>
</tr>
<tr>
<td>2</td>
<td>4.2 (0.7)</td>
<td>0.1 (1.2)</td>
<td>0.1 (1.7)</td>
</tr>
</tbody>
</table>

From these results, for Sub.1, the effect of bone-conducted sound on TTS at 4 kHz is not seen for the component of noise outside the critical band for it, but its effect is smaller even for noise of critical band than for noise of wider band.

For Sub.2, the greater effect of bone-conducted sound on TTS is caused by a wide band noise (425 - 3400 Hz), though there is no effect of sub-band noises. From this it may be said that the hearing protection by bone-conducted sound is primarily small for narrow band noise. To ensure it, we obtained the effect of bone-conducted sound on TTS caused by a pure tone of 30 Hz, which is the center frequency of the critical band (2400 - 3400 Hz), and could not find any significant effect.

From these experiments there may be a need to consider some hearing protection mechanism other than a middle-ear muscle contraction in relation to the effect of bone-conducted sound on TTS.

REFERENCE

1) Sone Toshio and Kono Shunichi, "The influence of subject's own voice in his noise dose," INTER-NOISE 81
2) Sone Toshio, Kono Shunichi and Tanaka Hiroshi, "The effect of the utterance of subjects in his daily noise exposure," INTER-NOISE 85
THE MEASUREMENT OF MEANING OF LOUDNESS, NOISINESS, AND ANNOYANCE IN DIFFERENT COUNTRIES

Seiichiro Namba*, Sonoko Kuwano* and August Schick**

* College of General Education, Osaka University
I-1 Machikaneyama, Toyonaka, Osaka, 560, Japan
** Institut zur Erforschung von Mensch-Umwelt-Beziehungen, Universität Oldenburg, Postfach 2503, D-2900 Oldenburg, F. R. Germany

1. INTRODUCTION

Over a number of years it has been suggested that it is desirable to differentiate loudness, noisiness and annoyance for the assessment of human responses to sounds.1 May,2 for example, makes the following differentiation:

"Loudness" is defined as the subjective magnitude of sounds. The unwantedness of the sound should not be considered. "Noisiness" is defined as the degree of unwantedness of a sound considered by itself. "Annoyance" is defined as the overall unwantedness of a sound heard in a natural situation.

In a natural or daily life situation, however, it is difficult to make clear differentiation of loudness, noisiness, and annoyance. The present authors and Thomas3 have measured the connotative meanings of the terms connected with noise problems in various countries, using semantic differential. According to the results, the semantic profiles of loudness, noisiness, and annoyance are very similar to each other in English. Also, the profiles of German loudness (Lautheit) and annoyance (Lastigkeit) are similar to each other. This means that both English and German "loudness" have negative connotations. On the other hand, Japanese "loudness" is quite neutral (Fig. 1-2). The profiles of English and Japanese "noisiness" were similar to that of "annoyance".

The results from this application of semantic differential suggest that the differences between the profiles of loudness, noisiness, and annoyance are not clear. Therefore, if it is accepted as being necessary to differentiate these terms for the assessment of noise, we need to reach general agreement on the appropriate definition of these terms.

In our previous research involving semantic differential, we did not use actual sounds, but only asked subjects to consider these terms. For the present study, we presented actual sounds and examined what kinds of adjective subjects used to express the impression of various sounds at various levels. In addition, we conducted a conventional semantic differential experiment using the same adjectives and the same sound stimuli.

2. EXPERIMENTS

2.1 Stimuli

Thirty-six stimuli were used in the experiment. Their level patterns are shown in Fig. 3. These stimuli, which were recorded on a FCM tape, had been used in our previous experiment.4

2.2 Adjectives

Adjectives used in the experiment are shown in Table 1.

2.3 Procedure

The stimuli were presented in random order through an amplifier (Yamaha CA XII) and four loudspeakers (Bose 101MM). Subjects marked adjectives which they thought appropriate for expressing their

Fig.1 loudness

Fig.2 annoyance

Fig.3
Table 1

| 大きい | loud | 小さい | soft |
| 嫌い | beautiful | 嫌い | strong |
| 重い | thick | 軽い | gruff |
| 痛い | hard | 柔らかい | pleasant |
| うるさい | ugly | すっきり | powerful |
| やさしい | gentle | 硬い | mild |
| 静か | quiet | 柔らかい | pleasing |
| 柔らかい | dull | すっきり | unsatisfactory |
| 適度な | moderate | 強い | sharp |
| 弱い | weak | かん高い | shrill |
| 騒々しい | clamorous | 聞こえがした | strident |
| 落ち着いた | calm | 好ましくない | unpleasing |
| 金属性の | metallic | やましい | noisy |
| うるさい | annoying |

Impression of each sound. They were allowed to mark as many adjectives as they wished. The 27 adjectives were presented in three sets, each with a different order.

2.4 Subjects

128 college students, aged about 18 or 19, were adopted as subjects.

3. RESULTS AND DISCUSSION

Part of the results are shown in Fig. 4 and Fig. 5. It is apparent that the frequency with which certain adjectives were chosen is different for different sound sources and different sound levels (Leq). Commonly, as the sound level decreased, the frequency of the choice of negative adjectives (such as "noisy" and "annoying") decreased. The frequency of the choice of adjectives connected with tone quality (such as "metallic" and "deep") was not noticeably affected by sound level. In the case of meaningful or useful sound sources such as speech and music, "loudness" was chosen as a suitable term for expressing the subjective strength of sounds. The same tendency was found in the other study in Germany. 3

The results of these experiments suggest that "loudness" is a different aspect of sounds from "noisiness" and "annoyance", and depends on the nature of the sound sources. Using this method, the difference between desirable and undesirable sounds in loudness judgments was expressed more clearly than when conventional semantic differential was used. Further analysis is necessary to investigate the connotations of adjectives used for describing sounds.

REFERENCES

EXPERIMENTAL RESULTS ON THE ANNOYANCE OF COMMUNITY NOISES AND GROUPING OF NOISE SOURCES

Minoru Sasaki

Department of Acoustic Design
Kyushu Institute of Design
4-9-1 Shiobaru Minamiku Fukuoka, Japan 815

Some experiments were made for rating the annoyance due to various kinds of community noises. The sound stimuli used in experiments are the noises which are often heard in our daily life.

In this report, the results of two previous experiments are used for considering the psychological effects with respect to the qualities of noise stimuli, and used for grouping of noise sources with respect to the content of information of stimuli by means of the factor analysis.

Experimental Procedure

The sketch of the test room is shown in Fig.1. The subjects are 17 males and 3 females who are from nineteen to twenty five years old. The jobs during noise exposure are the studying on mathematics in the degree of general course of college. The subjects are requested to rate on annoyance at a viewpoint of disturbing effects on the studying of mathematics. The spectrum equalizer is used for simulation of sound insulation characteristics through single mortar wall.

![Fig.1 Sketch of test room and block diagram of system](image)

Table 1 shows the sound stimuli used for experiment 1. The sound stimuli are seven kinds of community noises. These are the noises which are often heard in the residential area or the mixed area of residences and shops. The sound stimulus which is named "singing in café bar" is vocal singing accompanied by the tape recorded musical bands in café bar. This is the amusement of town people.

<table>
<thead>
<tr>
<th>No.</th>
<th>sound sources</th>
<th>L Aad</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>radio lesson of foreign language</td>
<td>74.2</td>
</tr>
<tr>
<td>2</td>
<td>barking of dog</td>
<td>60.8</td>
</tr>
<tr>
<td>3</td>
<td>singing in café bar</td>
<td>59.7</td>
</tr>
<tr>
<td>4</td>
<td>vocal training</td>
<td>66.6</td>
</tr>
<tr>
<td>5</td>
<td>piano training</td>
<td>61.6</td>
</tr>
<tr>
<td>6</td>
<td>game machine</td>
<td>65.3</td>
</tr>
<tr>
<td>7</td>
<td>singing in café bar</td>
<td>59.0</td>
</tr>
<tr>
<td>8</td>
<td>radio newreport</td>
<td>50.2</td>
</tr>
<tr>
<td>9</td>
<td>barking of dog</td>
<td>71.8</td>
</tr>
<tr>
<td>10</td>
<td>vocal training</td>
<td>72.6</td>
</tr>
<tr>
<td>11</td>
<td>piano training</td>
<td>70.6</td>
</tr>
<tr>
<td>12</td>
<td>radio newreport</td>
<td>61.2</td>
</tr>
</tbody>
</table>

However often causes of noise problems in Japanese town. The sound stimuli presented in experiment 2 are shown in Table 2. These stimuli are the superpositions of two or three kinds of noises shown in Table 1, and some of the transportation noises and voice of public address.

The A-weighted sound pressure levels are measured by the integrated sound level meter at the midpoint of the group of subjects. The L Aad values during the noise presentation are shown at the right column in Table 1 and 2.

Results of Two Experiments

The relationship between annoyance and L Aad values of the experiment 1 is shown in Fig.2. The correlation is comparatively low causing by diversities of the plots. Each point is an average of individual judgment values. In this research, these individual differences are considered with respect to the content of information of sound stimuli. It is supposed that the diversities of individual judgments are caused by the differences of disturbing effects due to the information which are included in the sound stimuli.

The results obtained in experiment 2 are shown in Fig.3. The correlation between annoyance and L Aad values is higher than in the case of the experiment 1. The standard deviation through all the plots is larger in experiment 1 than the case of the superposed sound of various kinds of noises, i.e., random or quasi random noise in the case of experiment 2. It is supposed that the annoyance of superposed noise is judged more uniformly among the subjects than the annoyance of noise from single source in a viewpoint of psychological disturbing.

![Fig.2 Relationship between annoyance and L Aad of each kind of community noise (after Table 1)](image)

![Fig.3 Relationship between annoyance and L Aad of superposed community noises (after Table 2)](image)
Effects, especially in the case when the sound stimuli contain the verbal or musical content.

Results of Factor Analysis and Grouping of Noise Stimuli

The sound stimuli used in the two experiments are factor analyzed using the annoyance rating values. The rotation of axes are calculated by the varimax method. The results of factor analysis of experiment 1 is shown in Fig. 4 (I-III axes). In this figure, the stimuli having large loading values on the I-axis are the radio news reports. These stimuli are the sound including the verbal content. The stimuli "singing in café bar" have large loading values on the III-axis. The stimuli "vocal training" has almost equal loading values on both the I and II axes. This stimulus is seems to be the complex type having the verbal and musical content.

The results of factor analysis of experiment 2 is shown in Fig. 5 and Fig. 6. In Fig. 5, the stimuli having large loading values on the I-axis are the sound including verbal and musical content. The stimuli having large loading values on the II-axis are the complex highly annoyed sound. In Fig. 6, the stimuli having large loading values on the IV-axis are the construction noises. These are the impulsive sound of concrete cutting. The stimuli having large loading values on the I-axis are the sound including the verbal and musical content.

From these results, the sound stimuli used in two experiments can be classified into four groups as follows:

- **group 1.** Sound composed of verbal content (verbal sound)
- **group 2.** Sound composed of musical content (musical sound)
- **group 3.** Sound composed of complex highly annoyed noise (complex annoyed sound)
- **group 4.** Sound composed of impulsive noise (impulsive sound)

Reference

ARE FLUCTUATING SOUNDS PARTICULARLY ANNOYING?
- AN EXPERIMENTAL STUDY

Peter Schaefer

Institut für Ergonomie, Barbarastrasse 16, D-8000 München 40

Introduction

Despite the wide variety of actual noise investigations, there are still basic shortcomings to be found in noise rating procedures (NRP). We feel that most of these flaws may be reduced to one central question:

Do these various NRP's yield good estimates for specific noise perceptions? Or more precisely, what is the performance of ratings like the L_a or the L_e (German "Beurteilungspiegel")? Do they produce good estimates even for long term perceptions? Unfortunately we mostly have to deny these questions.

This low appraisal means that there is only a narrow operating range for good performances of most NRP's - concerning in particular their actual noise spectra and their waveforms. That is, most NRP's are only working well if we are dealing with fairly constant sound levels and smooth broad band spectral forms. As soon as we are leaving this small range we are immediately facing serious problems.

What happens e.g. if we are changing to waveforms with varying characteristics, such as clearly oscillating sound levels? Are there shifts in our perception too? and if so, what is the response of the L_e, e.g.? - are there also corresponding shifts? Describing precisely the changes in noise perception? These are some of the most central questions we are dealing with in this paper.

Conception

In short we are starting at the following basic point:
- First we look at a sound clearly varying in time according to its sound level and - second we apply another sound with fairly constant characteristics;
- Finally we request test persons to adjust sounds regarding their annoyance respectively.

- The other sound - this is constant noise - will be adjusted by test persons as long as both sounds are assessed equally annoying.

Please note that we are referring to perceived annoyance.

This is easy said but harder realized experimentally. In practice we have to cope with a few particular problems:

1. So we know that the variety of different sounds is fairly unlimited. But here we decided for white noise (20 < f < 20 000 Hz) as basic signal because most every day sources show also wide frequency bands. Not least, the loudness of white noise in sones is easily calculated.

2. Next we have to generate appropriate oscillations (N_e(t)) of constant noise (N_e). We preferred full sine-modulations for nowadays its easy to decompose every signal into its harmonics. This modulation is realized by a simple multiplication of constant white noise (N_e) and a slightly modified sine-waveform:

\[ N_{e}(t) = N_{e} \left( 1 + \sin 2\pi f t \right) \]  \hspace{1cm} (1)

3. But which scale modulation should apply to as there are sound levels, sound power units or loudness measures in phones and sones? In our opinion we should only refer to scales that are good estimates of subjective noise perceptions (Schaefer, 1978). Of all measures mentioned above only perceived loudness in sones meets this requirement. Thus we refer to varying white noise signals (N_e(t)) that are fully sine-modulated regarding their loudness in sones acc. to (1).

4. Next question: what is the relevant frequency range of these modulations? We preferred the interval ranging from 0.1 to 50 Hz. The corresponding impressions may be associated to surf roaring, to purlings of a steam engine and to cracklings of a bale with raising frequencies. This range touches at its low end perceptions turning into single events and changes over to acoustic roughness beyond its upper end.

5. Finally we are to determine a fixed noise level for reference signals. For present we started at 70 dB.

This completes the listing of our most essential assumptions and we may proceed to the corresponding experiments:
- In a first experimental step we will tie up the modulated noise and the steady white noise will represent our test signal.
- In a second round roles will be reversed.

This means more precisely at the first step for example:
- We tune white noise on a fixed level at 70 dB.
- The loudness of this noise will then be fully sine-modulated acc. to (1).
- Modulation frequencies are selectable from 0.1 to 50 Hz.
- Finally test persons have to adjust levels of test sounds perceived equally annoying. Test signals are unmodulated white noises.

Fig. 1: Comparing fully modulated to constant white noise
Results

Results are presented in fig. 2. These are averages of 20 persons and 3 tests respectively; that is, every point represents 60 single values.

An easy access to interpretation begins in the middle of the chart. In this special case we referred to unmodulated white noise at a fixed level of 70 dB - this is the dotted line. Reference signals are generally filled up black in all 3 pictograms of the chart.

Test signals were of exactly the same kind and we noticed that they were all adjusted very close to 70 dB. Regarding this performance, we should take into account that adjusting homogenous sounds is a very easy task.

But the situation changes distinctly if we begin to modulate reference sounds as it is shown in the pictogram above:
- First the L of the reference signal rises to L = 74.2 dB due to modulation - this is the upper dotted line.
- On the other hand, test signals were adjusted to levels far above 74.2 dB varying with modulation frequencies - and this exactly is the effect we are about to trace out in this paper.

On the average these differences range between 8 to 11 dB with increasing modulation frequencies. So we realize that noise varying in time are much more annoying in human perception. In particular they are so much annoying that alternatively we are ready to accept an extra charge of more than 10 dB in constant noise. Low variations in noise levels are so estimated distinctly more comfortable. This is also the message of the lower line in the chart, bearing in mind the exchange of reference and test signals. Here we should remember once again that these results refer only
- to reference levels at 74.2 dB and
- to white noise as a particular spectral form.

So it remains to make sure, whether these results depend on variations of these two parameters. In any case more supporting research work must be done before we begin to integrate this effect in noise rating procedures. This integration might be done by a special weighting function based on the results in fig. 2. Weighting would apply to calculated loudness N(t) in sones. Here we notice a direct analogy to the well known A weighting procedure.

Conclusions

The message of this paper is shortly summed up as follows:
- We analysed time varying noise levels in view of their particular annoyance.
- To this purpose we compared modulated noises to constant noise.
- We found out that varying noises are considerably more annoying - so annoying that we are ready to spend more than 10 dB on constant noises, to get rid of noise oscillations.
- Finally there is a plausible way to integrate these results in existing NRPs.

Abbreviations

NRP Noise Rating Procedure
N(t) sone Loudness of constant white noise
N(t) sone Loudness of fully sin-modulated white noise in sones
L eq dB Equivalent sound level in dBs
L eq dB German "Beurteilungspegel" according to DIN 45 645
P(t) N/m² Effective sound pressure of constant white noise
P(t) N/m² Effective sound pressure of fully modulated white noise.

References


NOISE UNCOMFORTABLENESS WITH OTHER ENVIRONMENTAL FACTORS

By G. Horie, Y. Sakurai, T. Noguchi and N. Matsubara

Dept. Engineering, Kansai Univ., 3-3-35, Yamate-cho, Suita, Osaka 564, Japan, * Dept. Engineering, Mie Univ., 1515, Kaminakaicho, Tai, Mie, 514, Japan

Introduction

The "room condition" is composed of many different kinds of environmental factors such as the air temperature, the temperature of the walls, the air movement, humidity, the daylight coming through the windows, artificial lighting, noise level etc. Their negative effects on man can be characterized by the stresses they cause on man. Obviously, a common scale is needed for the evaluation of these stresses.

Since those personal judgements are generally not so much different from each other, in an ordinary room we judge the room condition to be not good or uncomfortable if such stresses are present. In such thermal conditions, the noise level and the lighting are combined in their moderate range, how do we evaluate their effects? If the thermal condition is changed, the uncomfortable changes depending only on it. And then if two other factors are changed, apparently the uncomfortable-ness they produce is added to that of the thermal condition. These empirical facts point to the existence of a simple rule of synthesis, such as the linear combination of the weighted effects of these environmental windrows.

The problem is different from the usual multi-variate analysis, because outside criteria are not quantitative but qualitative. It is possible to apply the Second Theory of Quantification of Hayashi to classify such a qualitative judgment by the weighted scores. In this paper, the outside criteria were classified on the mono-polar scale: 1) Neutral, 2) Slightly Uncomfortable and 3) Uncomfortable.

Second Theory of Quantification of Hayashi

When a synthesized subjective response of a subject i is $Y_i$, and the category weight score is expressed by $X_{jk}$, $Y_i$ is simulated to be linearly combined as,

$$Y_i = \sum_{j=1}^{r} \delta_i(jk) \cdot X_{jk} \quad (j=1,2,...; k=1,2,...)$$

where $\delta_i(jk)$ is a Kronecker's delta which is unity, if a subject i responds to the kth category of the jth factor, or else is zero. $X_{jk}$ belongs to the kth category of the jth factor out of r factors. A correlation ratio, $\eta^2$ is introduced as the measure which represents the degree of distinguishing the classes of the criteria. If the total variance of the outside criteria is $\sigma_Y^2$, the variance within a class is $\sigma_X^2$, and the variance between classes $\sigma_B^2$, the correlation ratio is defined by,

$$\eta^2 = \frac{\sigma_B^2}{\sigma_Y^2} = 1 - \frac{\sigma_X^2}{\sigma_Y^2}$$

with the relation: $\sigma_X^2 = \sigma_B^2 + \sigma_Y^2$. This ratio becomes larger and closer to unity, as the interval between the classes becomes larger, i.e. $\sigma_B^2$ becomes larger or $\sigma_Y^2$ becomes smaller. To get the largest distinction of classes, the ratio must be largest. Such weight scores $X_{jk}$ can be obtained by satisfying the following equations,

$$\sigma_Y^2/X_{jk} = 0, \text{ or } \sigma_B^2/X_{jk} = \eta^2 \sigma_Y^2/X_{jk} \quad (3)$$

Experiments

Four categories of thermal conditions which are different in summer and winter, four categories of traffic noise and three categories of illuminance were combined (see Table 1). One set of the experiments in a season consisted of 48 combined conditions. Each condition had, in principle, four subjects. Experiments were conducted four times a day in the anechoic room after its ceiling and walls were lined by wall papers, and the floor was covered by a plastic carpet. The relative humidities were 60±10% in summer and 40±10% in winter. The speed of air movement was less than 0.15 m/s at each subject during both seasons. The thermal radiation was measured with a globe thermomter at the center and was recognized to be almost equal to the air temperature near by. Traffic noise recorded along the highway was fed by two loud speakers in order to get a more uniform noise level distribution and the subjects were exposed to the noise just after entered in the room till they went out. The desks were illuminated by two 20 watts white fluorescent lamps and the illuminance on the desk top was controlled by the voltage supplied to the lamps. There were no general lighting. Physiological measurements, performance and questionnaires for the subjects during an experiment were scheduled as shown in Fig. 1. The subjects were prohibited to have a meal within two hours before the experiment and to have stimulants such as cigarettes and coffee within 30 minutes before the experiment. The "Uchida-Krappelis Test" which is a kind of an additive calculation of one figure, was adopted as a measure of performance.

Subjects were healthy male students whose ages were from 18 to 26. Hearing test were done to the subjects and all were approved normal. Clo values in summer were approximately estimated at 0.5, and those in winter were presumed to be about 1.0.

The Weight Scores of Each Category Given by the Second Theory of Quantification of Hayashi

The results obtained by the Second Theory of Quantification for three environmental factors with more than 500 samples over three years are shown in Table 1 for summer and winter. As the air temperature constitutes predominantly to the judgement of uncomfortable ness, the largest weight scores of the thermal conditions exists around 23°C. The response of uncomfortable ness is correlated to the noise level, but it does not change linearly. As for the illumination, its partial correlation coefficient is very small compared with other two factors. The partial correlation coefficient shows the degree of contribution of the factor to the outside criterion.

![Fig. 1 Schedule of the experiment](image_url)
Category weight scores in winter are listed in the
right of Table 1. The square root of the
reliability ratio was smaller than in summer. This
is probably caused by the smaller contribution of
room temperature, which is shown by the smaller
partial correlation coefficient. The contribution
of noise is closer to that of room temperature, and
that of illuminance is significantly smaller in winter.

Table 1. Category weight scores by the quantification.

<table>
<thead>
<tr>
<th></th>
<th>( \eta = 0.77 )</th>
<th>( \eta = 0.62 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>FACTOR ( \hat{\text{C}} ), ( \text{E} ), ( \text{P.C.G.} ), ( \text{E.C.G} )</td>
<td>( \text{W.E.} )</td>
<td>( \text{W.E.} )</td>
</tr>
<tr>
<td>( \text{N} )</td>
<td>( 60, 0.168 )</td>
<td>( 60, 0.143 )</td>
</tr>
<tr>
<td>( \text{Leq(A)} )</td>
<td>( 50, 0.151 )</td>
<td>( 50, 0.137 )</td>
</tr>
<tr>
<td>( \text{W.E.} )</td>
<td>( 70, 0.257 )</td>
<td>( 70, 0.257 )</td>
</tr>
<tr>
<td>( \text{W.E.} )</td>
<td>( 70, 0.175 )</td>
<td>( 70, 0.175 )</td>
</tr>
<tr>
<td>( T )</td>
<td>( 22, 0.012 )</td>
<td>( 22, 0.012 )</td>
</tr>
<tr>
<td>( \text{C}^\circ \text{C} )</td>
<td>( 30, 0.180 )</td>
<td>( 30, 0.180 )</td>
</tr>
<tr>
<td>( \text{C}^\circ \text{C} )</td>
<td>( 34, 0.137 )</td>
<td>( 34, 0.137 )</td>
</tr>
<tr>
<td>( \text{I} )</td>
<td>( 170, 0.015 )</td>
<td>( 170, 0.017 )</td>
</tr>
<tr>
<td>( \text{I} )</td>
<td>( 700, 0.065 )</td>
<td>( 700, 0.080 )</td>
</tr>
<tr>
<td>( \text{I} )</td>
<td>( 1480, 0.090 )</td>
<td>( 1480, 0.090 )</td>
</tr>
</tbody>
</table>

Fig. 2 Comparison of category weight scores in summer and winter.

**Prediction of A Room Condition by A Total Weight Score**

We can obtain the total weight scores for the
room conditions by summing the category weight
scores of three factors. Next, we have to find the
way to get a correct class by the total score with
the highest probability. When the probability
distribution of the subjective responses in each
class follows Gaussian, the dividing points \( Z_1 \), \( Z_2 \)
in Fig. 3 are obtained by the min-max method so that
the areas within each distribution curve equally
partitioned from \(-\infty \) to \( Z_1 \), from \( Z_1 \) to \( Z_2 \), and from \( Z_2 \)
to \(+\infty \). Referring to these two dividing points, we
can judge a room condition by three criteria.

Here are a few examples. If we have a noise
level 60 Leq(A), a room air temperature 30\(^\circ\)C and
illuminance on the desk of 170 lx in summer, the
category weight scores are 0.052, 0.180 and -0.057,
respectively and we get a total score of 0.175.
Referring to \( Z_1 \) and \( Z_2 \) in Fig. 3, we judge that this
environmental condition is slightly uncomfortable.

If we want the room condition to be neutral, the
total weight score should be more than 0.55. We
can discuss which kinds of condition can be changed.
If a noise level 60 Leq(A) and an illuminance 170 lx
are given, the weight score of room air temperature
must be more than 0.552. To obtain the neutral
condition, we can find that the room air temperature
has to be kept less than 27\(^\circ\)C, by the interpolation
between 26 and 30\(^\circ\)C.

Uncomfortableness As A Measure for Noise in A Room

The words used as a measure for noise are, for
example, annoyance, annoyance, etc. It is interesting
to see how uncomfortable noise used in this study
with those measures in the range of moderate room
noise. Unfortunately there have not been so many
studies on annoyance or annoyance in the range of 40
to 70 dB(A) or so. One of them, Young(1964) showed
that annoyance correlates virtually linearly with NC
or NCA, Lp and Perceived Noise level.

Uncomfortableness used in this study is
limited in the range from uncomfortable to neutral.
Since annoyance is a sensible scale, it is
natural that it correlates with decibel scale
linearly. Uncomfortableness is rather an emotional
scale than a sensible scale. Therefore,
uncomfortableness does not always parallels with
noise level but correlates to some extent.

Acknowledgement

The authors express their gratitude for the help
of Mr. Hiroshi Morimoto, the engineer at Kansai
University.

References

Hayashi, C., 1954, Multidimensional quantification
---with the application to analysis of social
phenomena---, Ann. Inst. Statist. Math., 5,
121-143.

Young, R.W., 1964, Single-number criteria for room
EXTRA-AUDITORY NOISE PROBLEMS IN EDUCATIONAL SETTINGS: STUDY RESULTS FROM DAY-CARE CENTRES FOR CHILDREN.

Claire Truchon-Gagnon et Raymond Héroux

Groupe d'Acoustique de l'Université de Montréal (École d'Orthophonie et d'Audologie) c.p.6128 succ. A, Montréal, Québec, Canada H3C 1A8

THE PROBLEM:

Settings which are noisy, but not noisy enough to cause hearing losses, have seldom been studied.

In educational settings, more specifically in day-care centres for children (DCC), what would be appropriate and safe noise exposure limits? Which building acoustical standards should be applied?

At the present time, English and French literature on noise do not provide guidelines for such settings.

In Québec, DCC workers and associations have expressed worries and complaints about noise exposure and noise-related problems, upon children and adults, in several DCCs.

An exploratory study has been undertaken to verify the following hypothesis, based on a preliminary inquiry: higher noise levels and more noise-related problems should be found in open-spaced DCCs, or in DCCs characterized by more noise transmission between groups of children.

METHODS:

A) Sampling: a sampling of 6 DCCs was designed accordingly to room acoustics criteria.

One DCC was "open-spaced", receiving all children (more than 40) in 1 large room, subdivided only by furniture and incomplete partitions.

2 DCCs were "partially open", subdivided in several (6-9) rooms separated by walls, some being acoustically inefficient.

The last 3 DCCs were subdivided in several (9-10) "closed-spaces" which were acoustically independent from each other: when doors were closed, one could not hear voices in adjoining rooms.

No significant exterior noise problem was encountered in any of the selected DCC.

B) Data Collection: the data consisted of a written questionnaire, addressed to the 105 workers of the 6 selected DCCs, and of systematic observations in one group of children, of similar age (the 2-3 years old), in each DCC.

The observations comprised the following:

1- room dimensions,
2- reverberation time,
3- number of individuals in the room,
4- number of interventions to limit noisy behavior among children,
5- noise measurements,
6- noisy events occurring during noise measurements.

C) Noise Measurements: a minimum of 700 noise measurements were performed, during a same year at winter-time, in each of the groups of children previously mentioned.

Half of the measurements consisted of L_Aeq-60sec, and the other half were L_pA max/60sec (highest peak value registered during an interval of 60 seconds).

These measures were performed alternately and consecutively.

To ensure a good representativity of the noise measurements, several criteria were adopted, including the following: the measurements were performed, for each group, during minimum sampling intervals of 4 half-days, on 4 different days: 2 whole morning-periods, from 9 a.m. to children’s nap, and 2 whole afternoon-periods, from the end of the nap to 6 p.m..

These periods were considered typical by the group’s workers and did not differ by more than 2 dB from each other (the a.m.-integrated L_Aeq being compared with a.m., and p.m. with p.m.) otherwise measurements were repeated until criteria were met.

During this study, 2 half-days have been considered as atypical and measurements were started afresh: the motives were, in one case, the presence of a new child who cried a lot, and, in the other case, the occurrence of a special event (a photographer’s visit) which changed the daily activity schedule and excited the children.

During all noise measurements intervals, noisy events (ex.: screaming, toys impacts, etc) were registered on a check list: when easily identified, the loudest event of an interval was also pointed out on that list, as well as all workers’ interventions about noisy behavior (and other details which might be relevant).

In one of the 6 DCCs, observations were performed over a larger number of periods: it comprised 8 half-days in a 2 months period, preceding the installation of sound baffles in the observed group’s room, and 8 other half-days in the following 2 months.

D) Questionnaire: a questionnaire was built after group interviews of approximately one hundred DCC workers. The final version was validated by 5 workers who, as those previously mentioned, did not take part in the inquiry.

The questionnaire consisted of 29 questions, mostly on a 6 points scales. It inquired about the following: noise sources and noise problems in the DCC, noise effects upon children and workers, and workers’ attitudes towards noise in general.

The questionnaires were distributed within 2 days in the 6DCCs, and the 93 completed forms (out of a possible total of 105) were picked up within the following 3 weeks.

RESULTS:

The major results consist of reported noise problems, noise levels and noise associated factors.

A) Noise Problems: In 2 of the sampled DCCs, ambient noise is described by vast majority of workers as being frequently loud and rarely comfortable, significantly more so than in the 4 other DCCs (t-test: p<.01).

In these 2 DCCs, health problems (such as fatigue, tension, headaches or other discomforts) are reported as being frequently or always present (during and after work-day) by most workers, as well as important work problems related to motivation, performance, satisfaction and communication (ex.: necessity of straining their voice to be understood, for 60%); all these effects are reported by a significantly higher proportion of these 2 DCC’s workers, than by others (t-test: p<.01).

Moreover, according to majority of workers of the 6 DCCs, noise has undesirable effects on children, including tensioness or restlessness and increased irritability or aggressivity. Some also report that noise incites children to behave more noisily. The problems reported as associated to noise, for children and adults in DCCs, are presented in Table 1.
TABLE 1. MAIN PROBLEMS ASSOCIATED TO NOISE IN DCC

<table>
<thead>
<tr>
<th>For children during work</th>
<th>For adults after work</th>
</tr>
</thead>
<tbody>
<tr>
<td>- restless or tenseness</td>
<td>- vocal efforts</td>
</tr>
<tr>
<td>- irritability or aggressiveness</td>
<td>- headaches or other discomforts</td>
</tr>
<tr>
<td>- tiredness &amp; tenseness</td>
<td>- tiredness</td>
</tr>
<tr>
<td>- communication problems</td>
<td>- voice disorders</td>
</tr>
<tr>
<td>- decrease in motivation, patience and satisfaction</td>
<td>- intolerance to noise</td>
</tr>
<tr>
<td>- headaches or other discomforts</td>
<td>- irritability</td>
</tr>
</tbody>
</table>

B) Noise levels: The measured noise levels, in the 6 groups of children, were high: no L_{Aeq-60sec} value was inferior to 61 dB (nap time excluded), while several L_{Aeq-60sec} values exceeded 85 dB everywhere. In each group L_{Aeq-8hrs} (integrating the a.m. & p.m. L_{Aeq-60sec} measured values) were found to be between 75 and 80 dB.

For each group, the median L_{Amax/60sec} fell between 100 and 107 dB; maximum L_{PA} levels registered varied between 117 and 120 dB. The differences between the DCCs noise levels are analyzed below.

C) Noise Associated Factors: highest noise levels were recorded while children were screaming or crying, playing with musical instruments (ex.: flutes) or during impacts (ex.: slamming a door: throwing or dropping a hard object, or hammering, on a hard surface).

Noisiest activity periods were those during which some of those noisy events could be heard (ex.: for a fanfare of 14 children, L_{Aeq-15min} was 89 dB; for a motor activity with hard vinyl hoops, with 7 children, L_{Aeq-23min} was 85 dB).

Other factors, likely to be associated to noise variations between DCCs, were studied through multiple regression analysis, in logarchymetrical & linear expressions.

Table 2. presents such analysis results, for periods of activities which were the most comparable, in the 6 selected groups: the first 20 minutes of meal (10 L_{Aeq-60sec/DCC}) and the first 16 minutes of story telling before nap (8 L_{Aeq-60sec/DCC}).

In Table 2, of all the variables studied, a few are presented which, in their logarithmic value, were found to explain a large part of the noise variance between groups, for such activities at least. These variables are: number of persons heard (N, varying from 11 to 44), the reverberation time in the room (TR, varying from 0.47 to 1.59 seconds), the room's total surface (TS, varying from 16 to 191m^2) and the workers' permisiveness (WP, varying from 1 to more than 10 daily interventions towards noisy behavior).

As seen on table 2, the number of persons heard (be it a larger number because of opened spaces, or because of inefficient acoustical partitions between rooms, or because of organization) explains more than 60% of the variance. Another part of the variance is explained, in one case, by room characteristics (TS & TR); in the other case, the reverberation time (TR) and permisiveness (WP) contributed significantly.

TABLE 2. VARIABLES ASSOCIATED TO VARIANCE OF NOISE LEVELS (L_{Aeq}) IN 6 DCCs, DURING MEAL TIME AND STORY TIME

<table>
<thead>
<tr>
<th>Activity</th>
<th>Variable % of predicted variance</th>
<th>cumul P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meal</td>
<td>Log N 68.2</td>
<td>68.2</td>
</tr>
<tr>
<td></td>
<td>Log TS 17.8</td>
<td>86.0</td>
</tr>
<tr>
<td></td>
<td>Log TK 5.2</td>
<td>91.2</td>
</tr>
<tr>
<td>Story</td>
<td>Log N 62.0</td>
<td>62.0</td>
</tr>
<tr>
<td></td>
<td>Log TR 23.2</td>
<td>85.3</td>
</tr>
<tr>
<td></td>
<td>Log WP 10.2</td>
<td>95.5</td>
</tr>
</tbody>
</table>

This corroborates results found when comparing workers' answers to questionnaire: the 2 DCCs where noise was described as the least comfortable, with significantly more noise related problems reported (see Results, section A) were:

1- the open-spaced DCC where more than 40 persons could be heard, all day, and 2- one of the 2 DCCs which had partially-opened rooms; in this case was found the highest number of rooms where some walls were inefficient acoustical barriers: at least 9 other persons could be heard at the same time as one's own group of 9 persons.

SOLUTIONS:

- in DCC for children, the number of individuals grouped in a room is a crucial factor: from a more detailed analysis of our results, number around 10 would be a maximum acceptable.

- Room acoustics is certainly critical in DCC, in both aspects of noise transmission & reverberation. It is worth mentioning that reducing TR from 1.6 to 0.6 seconds in one room reduced the group daily noise exposure by 5 dB, as well as the children's screaming frequency by nearly one half, even 2 months after sound baffles were installed. Other studies should also compare noise exposure in DCCs, before and after floor and furniture (hygienic but hard) surfaces would have been treated to absorb impacts.

- Educators can contribute significantly, but only after other factors have been controlled.

Acknowledgements: this study was conducted during the academic year of 1984-1985, thanks to a grant from Le Conseil des Universités du Québec (Fond de développement pédagogique/Volet des services aux collectivités)
A NEW DYNAMICAL ESTIMATION METHOD OF AN ENVIRONMENTAL
ACOUSTIC SYSTEM CONTAMINATED BY AN UNKNOWN NON-STATIONARY
BACKGROUND NOISE

Kazutaka HATAKEYAMA and Mitsuru OHATA

Faculty of Engineering, Hiroshima University,
Shintani, Sayo-cho, Higashi-Hiroshima City, 724 JAPAN.

INTRODUCTION

Generally speaking, in the field of environmental noise control, the instantaneous observations of random noise are contaminated by a background noise of arbitrary distribution type. And, to make matters worse, it is common in many actual measurement situations that an amplitude of background noise fluctuates randomly with the lapse of time and is usually unknown a priori.

As is well-known, a level fluctuation of environmental noise is brought on by various causes and exhibits an arbitrary distribution form of non-Gaussian type. Various evaluation procedures have been methodologically proposed from many different viewpoints, owing to the variety of phenomena and the complexity of the human response to them. In order to estimate or control precisely the actual environmental noise, some systematic methodology of removing the undesirable effect of background noise with unknown amplitude is definitely necessary from an engineering viewpoint.

Considering these facts, a unified dynamical estimation method based on Bayesian viewpoint is newly proposed in this paper, which can carry out simultaneously two kinds of estimations of unknown state of environmental noise energy system and the non-stationary fluctuating amplitude of random background noise. More concretely, by expanding a well-known Bayes' estimation theory into the hierarchical form under the introduction of a new measure of statistical independence between input signal and background noise, a computer-aided general algorithm of estimating both the unknown state and the amplitude of background noise based on the noisy observations is theoretically established in a recurrence form of wide sense digital filter matched to the actual successive observations.

Finally, the validity of proposed theory is experimentally confirmed by applying it to the actual noise data in a room acoustics.

FORMULATION OF ENVIRONMENTAL NOISE ENERGY SYSTEM UNDER
BACKGROUND NOISE WITH UNKNOWN AMPLITUDE

Let us consider the general system described by a non-linear time-variant dynamical model with a stationary random input \( w_n \) of arbitrary distribution type:

\[
x_{n+1} = \Phi_n(x_n, w_n),
\]

where \( x_n \) is the state variable at the \( n \)-th time stage. Furthermore, for expressing the successive observations of \( x_n \) under the existence of undesirable additive noise, we will formulate the physical measurement mechanism in a form of the following non-linear observation equation:

\[
y_n = \Phi_n(x_n, v_n) + w_n,
\]

where \( v_n \) denotes the non-stationary background noise process with a constant mean value \( v_n \) fluctuating with an unknown gain-factor \( b_n (v_n/v_n) \). The input \( w_n \) and background noise \( v_n \) are sampled independently along a time axis and are assumed to be independent of each other. The statistical information on the input \( w_n \) is assumed to be a priori given.

Now, let \( Y_n \) be the sequence of observations \( \{y_1, y_2, \ldots, y_n\} \). Given a realization of the observed data \( Y_n \), the discrete estimation problem consists of finding two kinds of estimates on arbitrary statistics of \( X_n \) and fluctuating unknown non-stationary gain-factor, \( b_n \), of \( v_n \) based on \( Y_n \). Then, consider Bayes' theorem on the conditional probability functions, as the fundamental relationship to derive an estimation of complicated phenomena:

\[
P(x_n | Y_n) = P(x_n | Y_{n-1}) P(y_n | x_n, Y_{n-1}) / P(y_n | Y_{n-1}),
\]

since the probability function on \( x_n \) can be derived by any statistical information on \( x_n \).

A UNIFIED ESTIMATION METHOD FOR ENVIRONMENTAL NOISE SYSTEM UNDER BACKGROUND NOISE OF UNKNOWN RANDOMLY FLUCTUATING AMPLITUDE

For the purpose of overcoming the difficulty of estimating the unknown signal \( x_n \) owing to an inevitable effect by the complexity of non-linear physical mechanism and the arbitrary non-Gaussian property of fluctuations, we will evaluate the nonlinearity of system and the non-Gaussian property by first introducing Bayes' theorem in a general expression form of orthogonal series expansion. From the analytical viewpoint of rapid convergence and the steadiness for the expansion type expression of Bayes' theorem, the appropriate type of probability density functions matched to the aim of study have to be selected as the first term of the series expansion. That is, these artificial probability functions had to be selected to describe the essential or fundamental probability characteristics commonly latent in the random phenomena under consideration. Then, by paying special attention to the conditional joint probability function \( P(x_n, y_n | Y_{n-1}) \) directly related to Bayes' theorem, we can obtain the orthogonally expanded expression of \( P(x_n, y_n | Y_{n-1}) \) in terms of the artificial probability functions on \( x_n \) and \( y_n \). \( P_n(x_n | Y_{n-1}) \) and \( P_n(y_n | Y_{n-1}) \), as follows:

\[
P(x_n, y_n | Y_{n-1}) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P_n(x_n | Y_{n-1}) P_n(y_n | Y_{n-1}) \Phi_n^{(1)}(x_n) \Phi_n^{(2)}(y_n),
\]

where \( \Phi_n^{(1)}(x_n) \) and \( \Phi_n^{(2)}(y_n) \) are orthogonal polynomials of degree \( m \) and \( n \) associated with the weighting functions \( P_n(x_n | Y_{n-1}) \) and \( P_n(y_n | Y_{n-1}) \), respectively. Thus, the following expression of Bayes' theorem can be newly derived in the series expansion form:

\[
P(x_n, y_n | Y_{n-1}) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P_n(x_n | Y_{n-1}) \Phi_n^{(1)}(x_n) \Phi_n^{(2)}(y_n),
\]

where each expansion expression \( A_n \) can be expressed as:

\[
A_n = \Phi_n^{(1)}(x_n) \Phi_n^{(2)}(y_n) = \int \Phi_n^{(1)}(x_n) \Phi_n^{(2)}(y_n) dx_n dy_n,
\]
in which the effects of unknown noise signal \( s_k \) and background noise \( v_n \) are explicitly reflected in the hierarchical form. It is noticeable that any statistics of \( s_k \) can be effectively estimated from Eq. (5) in the recurrence form matched to the computer treatment, if the unknown gain-factor \( b_n \) of background noise in Eq. (6) is determined in every repetition of estimations.

In order to identify \( b_n \) simultaneously with estimating \( s_k \), we will pay our special attention to the fact that the state variable \( s_k \) and the background noise \( v_n \) are statistically independent of each other. For the purpose of explicitly expressing the deviation from the statistical independency between \( s_k \) and \( v_n \), let us expand the joint probability function \( P(x_k, v_n | Y_{-1}) \) in an orthogonal series as follows:

\[
P(x_k, v_n | Y_{-1}) = P_x(x_k | Y_{-1})P_v(v_n) \sum_{s = 0}^{\infty} \sum_{b = 0}^{\infty} B_{sb} \cdot \phi^{s+1}(x_k) \cdot \phi^{s+3}(v_n),
\]

(7)

where \( \phi^{s+3}(v_n) \) denotes the orthogonal polynomial of degree \( s \) with a weighting function \( P_v(v_n) \), and the expansion coefficient \( B_{sb} \) is generally expressed as follows:

\[
B_{sb} = \left( \phi^{s+1}(x_k) \cdot \phi^{s+3}(v_n) | Y_{-1} \right).
\]

(8)

The information on various types of statistical correlations between \( x_k \) and \( v_n \) is reflected hierarchically in \( B_{sb} \) as the deviation from the product of the weighting functions \( P_x(x_k | Y_{-1}) \) and \( P_v(v_n) \). From Eq. (7), two marginal distribution functions \( P(x_k | Y_{-1}) \) and \( P(v_n) \) can be respectively expressed as follows:

\[
P(x_k | Y_{-1}) = P_x(x_k | Y_{-1}) \sum_{s = 0}^{\infty} \sum_{b = 0}^{\infty} B_{sb} \cdot \phi^{s+1}(x_k),
\]

(9)

\[
P(v_n) = P_v(v_n) \sum_{s = 0}^{\infty} B_{sb} \cdot \phi^{s+3}(v_n).
\]

(10)

Therefore, from Eqs. (8) and (10), the deviation \( \epsilon(x_k, v_n) \) expresses the statistical independency between \( x_k \) and \( v_n \) is given by:

\[
\epsilon(x_k, v_n) = P(x_k, v_n | Y_{-1}) = P_x(x_k | Y_{-1})P_v(v_n) \sum_{s = 0}^{\infty} \sum_{b = 0}^{\infty} B_{sb} \cdot \phi^{s+1}(x_k) \cdot \phi^{s+3}(v_n).
\]

(11)

When \( x_k \) and \( v_n \) are mutually independent, \( \epsilon(x_k, v_n) \) takes identically a zero value. So, by substituting the inverse relation \( (v_n | x_k, v_{-1}) = P(v_n | x_k, v_{-1}) = P(v_n | Y_{-1})P(v_n) / P_x(x_k | Y_{-1})P(v_n) \) into Eq. (11), the criterion function \( \epsilon(x_k, v_n) \) for the purpose of estimating the unknown gain-factor \( b_n \) so that \( x_k \) and \( v_n \) satisfy the property of the statistical independency.

Thus, substituting the relationship \( P(x_k, v_n | Y_{-1}) = P_x(x_k | Y_{-1})P(v_n) \epsilon(x_k, v_n) \) into Eq. (9), the following equality can be obtained:

\[
A_{mn} = A_{mn}^n + A_{mn}^v
\]

(12)

\[
A_{mn}^n = \sum_{s = 0}^{\infty} \sum_{b = 0}^{\infty} B_{sb} \phi^{s+1}(x_k) \cdot \phi^{s+3}(v_n) \int \frac{\partial \phi^{s+1}(x_k)}{\partial x_k} \frac{\partial \phi^{s+3}(v_n)}{\partial v_n} dxdv,
\]

\[
A_{mn}^v = \sum_{s = 0}^{\infty} \sum_{b = 0}^{\infty} B_{sb} \phi^{s+1}(x_k) \cdot \phi^{s+3}(v_n) \int \frac{\partial \phi^{s+1}(x_k)}{\partial x_k} \frac{\partial \phi^{s+3}(v_n)}{\partial v_n} dxdv.
\]

Eq. (13) shows that each expansion coefficient \( A_{mn} \) is composed of two terms: \( A_{mn}^n \) expresses the statistical independency between \( x_k \) and \( v_n \), and the term \( A_{mn}^v \) expresses the deviation from the statistical independency.

Then, based on the unified form of seriously expanded Bayes' theorem (5), the estimates of any kind of statistics given by an arbitrary polynomial function \( f_n(x_k) \) with an exponent less than \( N \) of state variable can be recursively obtained in the concrete expression form of orthogonal expansion series type, as follows:

\[
f_n(x_k, v_n | Y_{-1}) = \sum_{n = 0}^{N} \sum_{m = 0}^{\infty} (A_{mn}^n + A_{mn}^v) C_m \psi^{n+1}(y_m).
\]

(13)

where the coefficient \( C_m \) is determined in a way for the use of \( \psi^{n+1}(x_k) \).

Finally, the following unified algorithms of estimating the arbitrary statistics \( f_n(x_k) \) simultaneously with identifying \( b_n \) can be established, based on Eq. (13):

\[
f_n(x_k) = \sum_{n = 0}^{N} \sum_{m = 0}^{\infty} A_{mn}^n C_m \psi^{n+1}(y_m) / \sum_{m = 0}^{\infty} A_{mn}^n \psi^{n+1}(y_m),
\]

(14)

\[
= \sum_{m = 0}^{\infty} \sum_{n = 0}^{N} A_{mn}^v C_m \psi^{n+1}(y_m) = 0.
\]

(15)

It is noticeable that the above new results include some generalization on the type of estimation processes, viz., it can correspond to a general case when an arbitrarily distributed input noise passes through an arbitrary non-linear time-variant acoustic system, especially under random background noises with an unknown amplitude.

**EXPERIMENTAL CONSIDERATION**

We have estimated the sound level transmitted through the actual sound insulation system contaminated by a random background noise with unknown non-stationary amplitude. The experimental data are observed at the sampling interval of 0.2 sec. under the existence of telephone-bell sounds. One of the typical results is shown in Fig. 1.

Fig. 1 Estimated result of transmitted white noise in the reverberation room under the background noise.

**CONCLUSIONS**

The main purpose of this paper is focused on finding how to extend the well-known results by Kallman by generalizing Bayes' theorem in a new form of unified expansion, especially in order to be suitable for finding a recurrence algorithm on statistical quantities of arbitrary type of unknown signal, in the actual case when the amplitude of random background noise is unknown.

**REFERENCES**

PSYCHOACOUSTIC STUDY OF HUMAN RESPONSE TO TRANSMISSION LINE AUDIBLE NOISE

J.F.K. Foreman

Sound and Vibration Laboratory, Faculty of Engineering Science, The University of Western Ontario, London, Ontario, Canada, N6A 5B9

FIELD MEASUREMENTS

Field measurements were carried out in conjunction with the American Electric Power Service Corporation. Tests in Canada were conducted at sites adjacent to Ontario Hydro and Hydro-Quebec transmission lines and a Hydro-Quebec transmission station. The AEP tests were carried out at a site adjacent to an Ohio Power Company transmission line. Instrumentation, housed in trailers, was developed for automatic data logging (digital recording) of corona noise and associated environmental (weather) conditions; the laboratory also developed a microprocessor-based system for the automatic recording of corona sounds on four-channel analogue tape recorders (see Figure 1).

The analogue (sound) recording microprocessor system was programmed to insert a 90 dB (internally adjusted to 70 dB) calibration reference signal at the beginning of each channel on the tape. The microprocessor programming detected corona sound from the line (e.g., when the sound pressure level (SPL) in a 3% narrow band filter centred at 13 kHz equaled or exceeded 35 dB); the microprocessor was programmed to inhibit recording if the wind velocity exceeded 12 km/hr. When these requirements were met, the microprocessor then turned the tape recorder on for 2 minutes and off for 28 minutes, inserting, by means of an ACIA encoding system, the date, time, ID number of tape and station, and programming information at the beginning of each 2-minute recording period. On playback of the tape in the laboratory, this block coding was detected by a home-based microprocessor which gave exact date and time of recording. By accessing this same date and time in the digitally recorded data, a frequency identification was made of the SPL vs frequency (in octave bandwidths) of the corona sound, background noise, and weather conditions (wind velocity and direction, temperature, relative humidity, rainfall) associated with the sound recording on the analogue tape.

LABORATORY FACILITIES

The project involved playback of test analogue (sound) tapes which had been prepared from the field noise measurements (and which included representative samples of other environmental noise) (see Figure 2). The facilities of the laboratory, including a specially prepared listening room for subjective testing (which had been furnished and was acoustically calibrated), were used for the psychoacoustic testing. This listening room conformed to the 1/3 octave band pink noise frequency response specification (± 2 dB) for psychoacoustic tests stated in SAE AIE 1157 (1) over the frequency range from 63 Hz to 20 kHz.

PROCEDURE

A behavioural aversion procedure, the "paired comparison", was used to assess human aversiveness to noise. Reproduced samples of corona noise (four separate stimuli), transformer station noise, and other common environmental noises (jet engine, traffic, lawnmower, air conditioner) were compared with an artificial reference stimulus in the above-mentioned listening room. The artificial reference stimulus was an octave band of white noise centred at 1000 Hz. There were 32 participants (16 male, 16 female) evenly divided in the age groups 30 years and under and 45 years and over. The 9 test stimuli were presented to participants in random order. Each participant was audiometrically screened for hearing acumen. Each participant was involved in two separate test sessions, where the 9 test stimuli were presented in different order for each session. This resulted in 256 responses per stimulus.

During the listening tests, the participant had control of the sequential presentation of the test stimulus and the reference sound through annotated buttons on a console. A volume control knob on the console allowed the participant to adjust the level of the reference sound; by sequentially calling up the test stimulus and the reference sound and adjusting the volume of the reference sound, the participant then established a sound pressure level at which he/she judged the reference sound to be equally aversive to that of the test stimulus. This level was then recorded by the laboratory technician/operator.

DISCUSSIONS

Figure 3 shows a plot of primary annoyance settings (PAS) vs sound pressure level (SPL) for each of the nine test stimuli as averaged over responses; they are shown for the linear measurement scale (see reference 2 for similar plots for A-weighting and D-weighting scales). The PAS is the mean level (in dB) to which respondents adjusted the 1 kHz octave band reference sound for equal aversiveness to each stimulus. The SPL is the level at which each of the stimuli was presented. Also shown in Figure 3 is a least-squares fit line drawn through the corona points CRI to CRA. The linear extrapolation of PAS (aversiveness) with SPL (sound pressure level) can be justified on the basis of previous research (3,4).

The intersection (point 0) of the extension of the corona plots in Figure 6 with the PAS level of 50 dB (the PAS level for the traffic sound) indicates that for equal aversiveness to traffic noise presented at 58 dB Lin, corona noise would have to be presented at 46 dB Lin (i.e., 12 dB lower).

Figure 3 also shows that aversiveness to corona noise appears independent of corona spectra shape (CRI to CRA were each unique spectrum) and is directly related to sound pressure level of the noise. Figure 4 shows the ADB values (mean, standard deviation and range) for each of the test stimuli plotted for the linear measurement scale. The ADB values are the difference between the sound pressure levels at which the stimuli were presented, and the levels to which respondents adjusted the reference (1 kHz octave band of white noise), for equal aversiveness.

The ADB values for the A-weighting measurement scale are shown in Figure 5. (The procedures for transforming the ADB Lin values to ADBA, and similar plots for D-weighting, are given in reference 2.)

CONCLUSIONS

From Figure 4, it can be seen that: corona noise samples were equally aversive to the reference sound when they were 11 to 12 dB lower in sound pressure level; that was equally aversive to the reference sound when it was 11 dB lower in sound pressure level; and that traffic noise was equally aversive to the reference sound when it was 0.5 dB lower in sound pressure level.

When A-weighting was employed in the measurement system (Figure 5) the difference between the corona noise and traffic noise for equal aversiveness to the reference sound was reduced from 11.5 dB to approximately 2 dB. Thus when A-weighting is the
measurement system used, corona noise and traffic noise are roughly comparable in aversiveness.

When the mean values of the stimuli sound pressure levels relative to the reference sounds (the ADRs), together with standard deviations, were calculated for each noise measurement scale, A-weighting gives the smallest standard deviation (2.0 dB), the N-weighting the next smallest (2.5 dB), while the Lin-weighting gave the largest deviation (4.4 dB). (Pearson et al (5) and Molino et al (3) also made similar findings). Thus, if the criterion for "best noise weighting measure" is based on the least variability in subject responses to a range of stimuli, the A-weighting would appear to be the most acceptable for preliminary assessments, initial surveys and general applications involving measurements of environmental sounds (including corona noise).

REFERENCES


In developing a new procedure, the task team recognized five problems associated with the ANSI S1.13 method:
1. The procedure for measuring the masking noise level on either side of the tone is not clearly specified; different laboratories frequently obtain different results.
2. The procedure does not adequately reflect new knowledge of the psychoacoustical mechanism of masking.
3. The Fletcher critical bands should be replaced with new critical bands based on recent research.
4. The procedure should be extended to account for multiple tones within a single frequency band.
5. The procedure is time-consuming and difficult to automate.

The CBEMA task team undertook to develop a new tone identification procedure which overcomes the problems listed above. The new procedure may be used with FFT analyzers or other modern instrumentation such as programmable digital filters. The new procedure (i) computes the power in the tone; (ii) computes the total power of the Zwicker critical band centered on the tone; (iii) calculates the tone-to-noise ratio; and (iv) determines whether the tone meets the criterion for prominence, i.e., the tone-to-noise ratio equals or exceeds 6 dB. This criterion is the same as that used in the ANSI S1.13 procedure.

The new CBEMA procedure overcomes all of the problems associated with the ANSI S1.13 method. The technique for computing the tone-to-noise ratio is precisely defined and human judgment is not needed to decide where the two level readings are to be taken. Psychoacoustical research indicates that the noise power contributing to the masking of a tone is the total power contained in the critical band minus the power in the tone. The value of the total power is difficult to obtain using tunable analog filters, but it is easy to measure using digital techniques. The use of the narrower Fletcher critical bands in the ANSI S1.13 method ignored significant portions of the masking noise spectra. For this reason, the new procedure uses the Zwicker critical bands which are 2.5 times as wide as the Fletcher bands. The ANSI S1.13 procedure has no explicit technique for handling multiple tones within the critical band. In the new procedure, secondary tones are included in the total power in the Zwicker critical band that surrounds the primary tone. An important advantage of the new procedure is that it can be completely automated.

The American National Standards Institute has just approved a new standard, 4 ANSI S12.10-1985: "Methods for the Measurement and Designation of Noise Emitted by Computer and Business Equipment." Appendix B, "Identification of Prominent Discrete Tones" which includes the new procedure developed by the CBEMA task team. The procedure described in Appendix B does not require the use of any particular instrumentation; in fact, a variety of modern digital equipment may be used. Provided the accuracy constraints are met, nonetheless, it appears that the most accurate procedure is one which utilizes an
MEASUREMENT OF SOUND POWER AT HIGH FREQUENCIES

The measurement of noise emitted by computer and business equipment at frequencies above 10 kHz has posed a number of problems for the past two decades. Some equipment emits broad-band noise (e.g., paper noise associated with high-speed printing) while other equipment emits narrow-band noise and discrete tones (e.g., switching power supplies and video display units). In 1982, the European Computer Manufacturers Association (ECMA) initiated the development of a standard for measuring the high-frequency noise emitted by computer and business equipment. In 1984, ECMA established a task team, chaired by George C. Maling, Jr., of IBM Poughkeepsie, which was instructed to work with ECMA in developing a suitable standard. The two trade associations have now completed their task with the approval by the ECMA General Assembly of a new ECMA standard on this subject.

The ECMA standard describes four methods for determining the sound power levels of sources in the frequency range covered by the octave band centered at 16 kHz. The first three methods are intended for use in a reverberation room, and the fourth method is carried out in a free field over a reflecting plane (hemi-anechoic chamber). The sound power levels so determined are typically subject to standard deviations of approximately 2.5 dB.

The three reverberation room techniques require a determination of the average sound pressure level produced by the source in the reverberant field, \( L_p \). In the first method, the room constant \( R \) is calculated from the measured reverberation time using the Eyring equation. The sound power level is calculated from the well-known equation:

\[
L_w = L_p - 10 \log \frac{R}{R}
\]

The second method is a simplification of the above technique which avoids a measurement of the reverberation time. The room absorption is calculated directly from the air absorption coefficient, and the sound power level is calculated from the same equation as above.

The third method uses a reference sound source (RSS) which emits sufficient acoustical energy in the 16 kHz octave band to obtain an average band pressure level in the reverberation room which is at least 10 dB above the background noise level. The sound power level is then calculated using the comparison equation:

\[
L_w = L_{w(RSS)} - L_{p(RSS)} + L_p
\]

The fourth method uses a free field over a reflecting plane. Although air absorption plays an important role in the high-frequency range, its effect is negligible for a measurement radius less than 0.1 m. The sound pressure level is measured at individual microphone locations or on prescribed microphone paths, the surface sound pressure level, \( L_{pf} \), is determined. The sound power level is then calculated from the surface sound pressure level and the area of the measurement surface, \( S \):

\[
L_w = L_{pf} + 10 \log \frac{S}{S_0}
\]

where \( S_0 \) equals one square meter.

ECMA has recently submitted the new standard to the ISO for conversion to an International Standard. Working Group 23 has recommended that this conversion proceed as quickly as possible, thus ensuring compatibility between industry and international standards.

MEASUREMENT AND EVALUATION OF IMPULSIVE NOISE

The third task team established by CBEMA in 1984, chaired by Robert Hallweg of Digital Equipment Corporation, was instructed to clarify the procedures for evaluating and reporting the impulseliveness of a sound source. The results of this study were incorporated into the comments of the U.S. Member Body on ISO DIS 7779, which were submitted at the end of 1985. The finalizing procedure is recommended: after aural examination indicates that impulsive noise is present, the difference in decibels between the A-weighted impulse sound pressure level and the A-weighted sound pressure level is obtained. If this difference is equal to or greater than 3 decibels, the noise is identified as being impulsive.

Through close cooperation between the international standards body (ISO), the trade associations (CBEMA and ECMA), and national standards organizations which are concerned with noise emission standards, a test code has been developed for the computer and business equipment industry which is acceptable to all parties concerned.

REFERENCES

THE REDUCTION OF STRUCTURE-BORNE NOISE IN TRANSFER MACHINES

T.N. Moore
Queen's University
Department of Mechanical Engineering
Kingston, Ontario, Canada
K7L 3N6

INTRODUCTION

In recent years significant progress has been made in reducing noise levels generated during the operation of high volume, multitasking transfer machines. Yet, in spite of these improvements, there still exists the need for further decreases in machine noise levels.

The noise control methodology employed up to this point in time has consisted of a two-pronged approach. Firstly, where possible, the noise is reduced at source by identifying the noise generating mechanisms and then modifying them to achieve quieter operation [1]. Secondly, when this approach is not practical or does not provide sufficient reduction, noise barriers and enclosures are employed to provide the necessary attenuation [2].

However, it has become increasingly clear that further significant noise reductions "at source" are not likely in the short term. Also, the use of barriers and enclosures has already reached the point where any further improvements in noise reduction could only be achieved by degrading the production efficiency of the machine.

Transfer machines consist of a large number of workstations attached to a central transfer spine. The workpieces are transferred sequentially into each workstation for machining (milling, drilling, boring, etc.). A given workstation consists of a machining "module" (which carries the cutting tools) and a slide unit assembly (which cycles the head's cutting tools into and out of the workplace) all of which are supported by a "wingbase". See Figure 1. For simplicity of discussion the combined head unit and slide unit assembly will subsequently be referred to as the "head assembly".

The wingbase is a boxlike (open at the bottom) steel weldment which is bolted both to the central spine and also to the machine's foundation. The head assemblies at each workstation are significant generators of structural vibration (due to excitation from gears, timing belts, bearing elements, metalcutting operations, etc.) which is transferred to the wingbases. The large flat surfaces of the wingbases are potentially efficient radiators of noise from this structure-borne vibration.

Recognizing that this "indirect" source of noise could, in fact, be a significant part of the overall machine noise level, a study was initiated to quantify the typical noise contribution of a wingbase radiating structure-borne vibrational energy. In addition, if the noise was indeed found to be of significant magnitude, then possible structural modifications to reduce contributions from this source were to be investigated.

METHODOLOGY AND RESULTS

quantification of the Wingbase Noise Amplification Effect

The first step in the study was to determine the noise amplification effect produced by a wingbase. This was accomplished using a test rig which consisted of a typical two-spindle drill head assembly and its associated wingbase as shown in Figure 1. All measurements in this study were obtained with the spindles turning but without metal cutting. The noise measurements were made at various positions around this rig, each at a distance of 90 cm from the nearest radiating surface and at a height of 122 cm from the floor.

Noise measurements were made for two cases. First, the head assembly was suspended 2.5 cm above the wingbase with no structural connection between the two units. This maintained their normal geometrical relationship without any structural coupling. In the second case, the head assembly was bolted directly to the wingbase. This is the normal operating configuration. Therefore any increase in noise level measured in the second case would be a result of wingbase noise radiation due to structural vibration.

It was found that the noise level did indeed increase on the order of 5 dB as a result of the structural coupling. Since this was a significant increase, several methods of reducing this structurally radiated noise were investigated.

Reduction of Structurally Radiated Noise

Generally speaking, isolating the head assembly from the wingbase using elastic elements was not expected to be a practical proposition due to the requirement for extremely accurate positioning of the head during machining. Thus, attention was focused on structural modifications to the wingbase in an effort to make it a less efficient noise radiator.

To begin, both noise and vibration (acceleration) data were obtained from the drill head assembly and analysed in both the time and frequency domain. This information was obtained prior to mounting of the head assembly to the wingbase. After bolting the head to the wingbase, the noise measurements were repeated and vibration data was obtained from accelerometers mounted at various locations on the wingbase. This procedure allowed the structure-borne vibration which produced the most significant noise components to be determined.

Figure 2 shows a comparison of the typical noise spectra obtained for the head assembly suspended in air above the wingbase and for the head assembly bolted directly to the wingbase. The major contribution of spectral components in the range of 600 to 1800 Hz is quite apparent. Note also the significant increase in level over this range as a result of the structurally radiated noise from the wingbase.

Figure 3 presents a comparison of the acceleration spectra obtained for two measuring configurations. The first was with the head assembly suspended in air prior to being bolted to the wingbase. In this case, the accelerometer was mounted in the vertical direction on the head assembly mounting flange (i.e., at the normal head assembly/wingbase interface). Thus, this acceleration spectrum represents the typical structural vibration input to the wingbase.

In the second configuration the head assembly was bolted to the wingbase. The figure shows a typical acceleration spectrum obtained from the wingbase itself after the head assembly had been bolted to it. It is apparent that predominant components of the output coincide with predominant components of the forcing input.

Modal Analysis

To help determine which structural modifications would be most effective in reducing the vibration response (and thus noise radiation) of the wingbase, a modal analysis was performed.
The modal analysis was performed using the "impact method" [3]. In this procedure the structure is excited by a hammer blow. The resulting impact force is measured with a force transducer mounted in the head of the hammer. The vibration response is measured at various points on the structure using an accelerometer. This data permits the calculation of a Frequency Response Function (FRF) which is the ratio of the spectral content of the output signal divided by the spectral content of the input signal, over the required frequency range.

These measurements are repeated for each point of interest on the structure. Obviously, care must be taken to be sure that such a point is not a node (a point of no motion) for a mode of interest.

A review of the relevant modes indicated highly complex shapes which varied considerably from mode to mode.

As a result of this analysis, it was apparent that the incorporation of a "reasonable number" (considering costs of fabrication and materials) of stiffeners within the wingbase would not produce a significant vibration reduction. Since the modes of interest have complex shapes which change significantly from mode-to-mode, an excellent stiffening location in one mode is a poor choice in another. Thus "point-to-point" stiffeners would not be expected to produce significant vibration reduction.

The use of additional damping is another means of reducing vibrational energy and its associated noise generation. Localized damping pads at specific positions on the wingbase were not expected to be particularly effective since damping is most effective at points of large bending deflections, but these locations change significantly for the various modes of interest.

However, in contrast to stiffening, increasing damping over the whole structure is an economically justifiable solution. In this case, damping paste can be applied to the complete inside surface of the wingbase for little additional cost.

To evaluate the efficacy of such a solution, the complete inside surface of the test wingbase was coated with a 0.95 cm thick layer of a high quality damping paste. Subsequent vibration measurements showed a significant reduction in acceleration levels. While the noise measurements showed a typical reduction in overall noise level of 4 dB.

CONCLUSIONS

i) It has been shown that the wingbase elements of transfer machines are significant radiators of structure-borne noise.

ii) Using signature analysis methods, the significant components of this structure-borne noise were identified and targeted for reduction.

iii) Modal Analysis permitted various possible vibration (and hence noise) reduction strategies to be assessed without requiring the use of costly trial and error methods.

iv) The use of damping paste on the wingbase structure was identified as a cost effective method of reducing wingbase vibration. Typical noise reductions of 4 dB were achieved.

REFERENCES


EXHAUST NOISE CONTRIBUTION BY THE SCAVENGING FAN OF A HIGH EFFICIENCY DOMESTIC FURNACE

S.E. Brackett, V. Berends and R. Gaspar

Department of Mechanical Engineering, University of Windsor, Windsor, Ontario, Canada, N9B 3P4

INTRODUCTION

The domestic furnace industry has recently brought forth energy efficient furnaces. The design changes have resulted in furnaces which are approximately 50 percent efficient. These designs have gained popularity because of their economic benefit but have also introduced possible environmental noise problems. There are several potential sources of noise from these new furnaces which include scavenging noise (fan or pulsed pressure), and combustion noise. This is a situation which is of concern to the Noise Control Branch of the Ontario Ministry of the Environment. This is because many new residential subdivisions are being developed in which all the housing will use these furnaces. One of these manufacturers, Domatic Olsen Inc., in collaboration with the Department of Mechanical Engineering at the University of Windsor, is involved in research which is presently focused on minimizing the noise caused by the scavenging fan associated with their HCS 80/90 furnace.

EQUIPMENT

The domestic furnace that has been studied is the HCS 80/90 model manufactured by Domatic Olsen. The induced draft centrifugal blower unit of this furnace scavenges the combustion gases by inducing a negative pressure difference between the outlet and inlet of the combustion chamber. The minimum permissible pressure drop for continuous furnace operation is a pressure difference of 15.2 mm (0.6 inches) of HgO; however, the actual pressure operating drop should exceed 20 mm HgO (0.8 inches). An ASHRAE Fan Test was conducted on the original scavenging fan and furnace to determine the operating air flow rate [1]. Having established the pressure and flow rate requirements of the blower unit, noise measurements were made. This provided a base line against which modifications to the rotor of the scavenging fan could be evaluated.

SCAVENGING FAN NOISE EVALUATION

The noise generated by the scavenging fan is of a broad band nature with discrete frequency components [8]. This type of noise can be divided into two distinct aerodynamic classifications. These are rotational and vortex noise.

Rotational noise is characterized by the superposition of discrete tones of the blade passing frequency and its higher harmonics. These originate from two sources. The first source is the impulse that is given to the air each time a blade passes a given point [2]. The impulse consists of a steady component which provides the pressure drop across the fan and very small oscillating components which produce the rotational noise at the discrete tones. The second source is the result of each blade passing the cutoff point in the scroll cage where pressure pulsations occur.

The primary method of reducing rotational noise is to change the cutoff clearance. Design guidelines show that fan efficiency approaches a maximum when the cutoff clearance is a minimum. This is because almost all the air flow is discharged out of the fan and a minimal amount of air is recirculated. However, rotational noise is inversely proportional to the cutoff clearance, so that the minimum cutoff clearance causes maximum noise. The normal compromise accepted by manufacturers is a cutoff clearance of 5 to 10 percent of the rotor diameter. A clearance of 5 percent is considered to be the allowable minimum. At this clearance, the changes in fan efficiency are negligible. When noise control is a significant factor, a cutoff clearance of 10 percent or greater is considered desirable.

Vortex noise results from the shedding of vortices (eddies) from either the leading edge or trailing edge of the rotor blades. The noise generation from eddies is caused by the separation of the laminar sublayer of air flow over the blades which result in pressure fluctuations along the blade length. This type of noise is inherently more difficult to eliminate by ensuring a continuous laminar sublayer of air flow along the rotor blade and can be achieved by designing a proper blade profile.

SCAVENGING FAN BLADE SELECTION

The selection of the type of scavenging fan is based on a number of fan design parameters. The most important of these is the design fan speed. The original scavenging fan is of a forward-curved blade profile with 48 evenly spaced blades. The combination of blade profile and rotational speed has a significant effect on the noise generated. The influence of blade design and rotational speed on the noise generated by a centrifugal fan must be considered.

Centrifugal fans can be subdivided into three different categories depending upon the curvature of the rotor blades. They are as follows:

1. Backward-Curved-Blade Fans
2. Forward-Curved-Blade Fans
3. Radial-Blade Fans

The radial-blade fan is not as common as the forward-curved-blade and the backward-curved-blade fan and hence, it will not be discussed in this paper. The selection of fan type depends upon the operating speed for a required output. For a given output, a forward inclination of the blade should be operated at a relatively low speed, and a fan of backward inclination should be operated at a relatively high speed [7]. The noise characteristics of each type will now be discussed.

Forward-Curved-Blade Fans

As mentioned above this type of fan is characterized by the forward inclination of its fan blades and a lower operating speed as shown in Figure 1. The number of fan blades associated with this type of fan varies from 36 to 64 [6] or 32 to 64 [5]. A tip angle which is greater than 90° is normal for this blade type. Hence, the blade passing frequency for this fan type is higher than that of the other fan types. The fan blades of the forward-curved fan are usually shallow in depth; consequently, there must be more blades to provide the proper guiding influence of air flow through the channel between adjacent blades [6].

The velocity triangle for the forward-curved-blade fan is illustrated in Figure 1. The distinguishing characteristic for this fan type is the absolute velocity entering the scroll V2 REL. The tangential velocity, V2 REL, which increases because of the blade shape. The shape of the blade is such that the relative air velocity leaving the blade passage, V2 REL, and the tangential velocity, V2 REL, add vectorially producing an increase in the air velocity. This tends to cause the pressure to rise and an increase in pressure results in an increase in the noise level.

A forward-curved-blade fan is used in situations where the operating speed is relatively low. This operating speed tends to minimize any motor bearing
Backward-Curved-Blade Fans

This type of fan is characterized by the blades being backwardly inclined with the tip angle being less than 90° and having a relatively high speed of operation. The blade passing frequency for this type of fan is lower than that of the forward-curved-blade design because of the number of blades used. The number of blades recommended for a fan blade of uniform thickness varies from 12 to 16 [5] or 10 to 16 [6]. This provides a proper channeling effect of the air flow. In smaller sized fans, the number of blades may be as few as 8 and in the larger sizes as many as 24 [6].

The velocity triangle for the backward-curved-blade fan is illustrated in Figure 1. The absolute velocity entering the scroll, \( V_{2\,\text{ABS}} \), is less than the tip speed for this type of fan because of the blade shape. The relative velocity of the air leaving the blade passage, \( V_{2\,\text{REL}} \), is in the opposite direction as that of the tangential velocity, \( U_2 \). Therefore, when vectorially summed, this results in a decrease in the air velocity as compared to the forward-curved-blade. This in turn causes a decrease in pressure and a reduction in the noise level.

Although the noise generated by the fan is reduced by implementing backward-curved blades, the relatively high speed of operation may create another noise problem. In designing such a fan unit, care must be taken in choosing the proper bearings. Due to the high speed of operation associated with backward-curved-blade fans bearing noise can increase.

PROTOTYPE FAN DESIGN

Considering the above factors, including the design fundamentals, a computer program was developed. Of the computer based designs which were generated, two designs were chosen for prototype construction. These designs were based on the control of the pressure coefficient, as suggested by Eck [3], which should provide a minimal fan noise contribution.

The first prototype fan was designed for the minimum permissible pressure drop through the furnace. The entrance angle and the tip angle were set at 35° and 45° respectively. For this 12 bladed fan design the inlet breadth was calculated to be 1.3 cm, with the outlet breadth being 0.4 cm. However, for ease of construction the breadth of the inlet and outlet was set at 1.3 cm. The pressure coefficient resulting from calculations was 0.92. This prototype produced a volumetric flow rate of 0.81 m³/min (28.5 cfm).

The second prototype was constructed for the operating volumetric flow rate of 1.11 m³/min (39.2 cfm). This design was far more critical than the first because of the furnace design constraints. This volumetric flow rate is more conservative and provides approximately 180% more air than that required for complete combustion. The second prototype produced the required volumetric flow rate and had a corresponding pressure drop of 38.6 mm H₂O (1.52 inches) of water. The 16 bladed fan design had a pressure coefficient of 1.21 with a 10° entrance angle and a 45° tip angle. The breadth of the scavenging fan was held constant to the inner breadth of 2.3 cm.

RESULTS

Sound pressure levels were measured at an angle of 90° and a distance of one metre from the centerline of the end of the exhaust pipe. The 'stock' fan produced a sound level of 66 dB(A) while the first prototype produced a sound level of 60 dB(A). The second prototype produced a sound level of 61 dB(A). Therefore, the second prototype produced a sound level 5 dB lower than the original fan.

CONCLUSIONS

The preliminary results show that the implementation of a backward-curved-fan design can not only match the performance of the original scavenging fan but can also lessen the contribution of the exhaust noise to the surrounding environment. As the magnitude of the scavenging fan noise was about the same as the combustion noise contribution, the net result of the present work is a reduction of 2 dB in the overall sound level generated by the furnace. Additional work is presently being carried out in an attempt to control the combustion noise contribution.

REFERENCES


ACKNOWLEDGEMENTS

The authors of this paper would like to thank Mr. B. Berenda (Toy Tools Inc.) who machined the prototypes. Without his contribution this research could not have been completed. The authors are grateful to Duomatic Olsen for their contribution in supplying the HCS 80/90 high efficiency domestic furnace and induced draft blower units. The authors would like to acknowledge the contributions of Mr. R. Tattersall who provided technological support and Professor W.O. Colborne who provided input with respect to fan principles and furnace operation.

![Diagram of fan discharge](attachment:image.png)

**Figure 1**
FLUID-BORNE NOISE TRANSMISSION THROUGH MULTI-TUBE HEAT EXCHANGERS

N. Purshouse

YARD Ltd., Charing Cross Tower, Glasgow G2 4PP. United Kingdom

Fluid flowing through the pipework of a cooling system frequently has to negotiate passage through a heat exchanger which consists of a bank of small diameter tubes arranged in a regular array. Not only does the tube bank constitute a resistance of possible hydrodynamic significance, it can also have a bearing on the acoustics of the fluid system of which it is a part, influencing the reverberant sound level generated by an acoustic source positioned in the fluid.

In conventional heat exchangers the tubes are straight and all have the same length (Fig. 1). For acoustical purposes that type of tube bank may be regarded to a good approximation as a constriction of the main conduit: the transmission and reflection coefficients to incident plane waves are essentially those of a 'reverse' expansion chamber silencer of the type described by Dowling and Frowes Williams (1983). For although the sectional area of each individual tube in the array may be very small compared with that of the main pipework, an incident plane wave drives the wave motions in all the tubes in unison with one another, and the tube bank then behaves almost as if it were a single tube with sectional area equal to the summed areas of all the tubes in the array. Well known results then identify the principal acoustical parameters as the overall contraction ratio and length, the transmission loss taking maximum and minimum values when the array length bears certain fixed relationships to the sound wavelength.

Very different are some novel heat exchangers which are made up from tubes of many different lengths, each tube shaped into the form of a 'U' (Fig. 2). In terms of hydrodynamic flow resistance the performance of this type of heat exchanger is almost indistinguishable from that of its conventional counterpart, but the U-tube array has its own distinctive acoustical properties that will be examined in this paper.

In typical heat exchangers the tube diameters are small in comparison with the sound wavelength at most frequencies of interest, so that any curvature of the tubes is of no special acoustical significance. Much more relevant are the different tube lengths represented in the array, for waves triggered simultaneously at one side of the tube bank, but which travel through it in tubes of different length, clearly cannot arrive together at the far side. The phase differences which arise in this way admit more subtle inter-tube interactions than are possible when all the tubes have the same length, with significant implications for the acoustics of the tube bank as a whole.

The behaviour of an array such as this is examined in the paper with the aid of the theory of one-dimensional wave propagation in branched piping systems, described by Lighthill (1978). Viscous effects are found to be important under certain circumstances. Formulae for the transmission and reflection coefficients of typical arrays will be given in the paper, together with some approximate results valid in relevant asymptotic limits.

This work was carried out under contract to the Procurement Executive, Ministry of Defence, whose support is gratefully acknowledged.

© Crown copyright

REFERENCES

(1) Dowling, A.P. and Frowes Williams, J.E. 1983 Sound and Sources of Sound. Chichester: Ellis Horwood.

CAVITATION NOISE AND VIBRATION CONTROL IN DIESEL ENGINE IN-LINE FUEL INJECTION SYSTEMS

S. D. Haddad

College of Engineering and Applied Sciences
Western Michigan University, Kalamazoo, MI 49008

INTRODUCTION

Diesel engine fuel injection systems of the rotary and In-line jerk types as well as the P.T. unit injector type normally produce noise and vibration levels lower than engine combustion excitation, piston slap, timing gear rattle and bearing impacts. However, at certain design and operating conditions, these F.I.E. can produce high frequency excitations which induce fuel cavitation coupled with erosion damage in the various parts of the fuel system. Initial investigations using specially constructed motoring rigs have shown that the P.T. and rotary F.I.E. are less prone to this cavitation phenomenon than the In-line F.I.E. This paper presents results of a study conducted on the latter using an experimental-analytical approach.

EXPERIMENTAL STUDY

Haddad and Russell (Ref. 1, 2) have been able to construct special motoring rigs to study the noise and vibration characteristics of F.I.E. in isolation of other diesel engine noise and vibration sources. Figure 1 shows typical noise versus speed relationship of P.T. systems (Ref. 1) and In-line F.I.E. (Ref. 2) compared with overall noise of a low noise diesel engine developed by Haddad and Priebe during the period 1970-76.

![Graph of measured cavitating InLine F.I.E. noise levels over engine speed]

Figure 1 - Noise of motored In-line and P.T. fuel injection equipment (full fueling) compared with a low noise diesel engine (full load) and a cavitating In-line F.I.E.

- Overall engine and F.I.E. noise levels were measured at 1M from engine and F.I.E. outer surfaces respectively.
- Cavitation noise in the In-line F.I.E. was confirmed using an ultrasonic detector (1984).

![Diagram of hydraulic and gas pressure on tappet and piston]

**Figure 2 - Sketch to show the analogy between engine piston/liner (piston slap) behavior and In-line F.I.E. pump tappet/bore (tappet slap).**

It can be seen that the P.T. system is inherently quieter, while the In-line system is basically noisier, especially at higher engine speeds. In 1984, it was also possible to observe and measure excessive cavitation noise from an In-line F.I.E. to exceed engine noise as shown in Figure 1. Analysis has shown that the generation of tappet slaps are the main cause of this phenomenon.

ANALYTICAL STUDY

Tappet slap in In-line fuel pumps is an impulsive type mechanical excitation which could be modeled theoretically for parametric evaluation of noise and vibration reduction. Figure 2 shows the similarity between the problem of engine piston slap and that of In-line fuel pump tappet slap.

Haddad and Howard (Ref. 3) have developed a computer program to optimize piston/liner designs and operating conditions for low noise diesel engines. This program was modified to solve the problem of tappet slap in In-line F.I.E. shown in Figure 2 (Ref. 4). The tappet slap program has been used to predict the effect of tappet/bore clearances as shown in Figure 3 and other parameters. The program predicts tappet/bore behavior and calculates the kinetic energy lost at major tappet impact as well as the total kinetic energy lost per engine cycle. Since the contribution of potential energy at impacts is negligible, it was found that the total kinetic energy per cycle is approximately proportional to tappet bore vibration levels which explains the relationship between the severity of F.I.E. vibration levels and tappet/bore clearances shown in Figure 3.
Major tappet impact forces were calculated as follows:

\[ \begin{align*}
M \dot{x} & = F = K \Delta T \\
M x & = \frac{K (\Delta T)^2}{2} + C_1 \\
M x & = \frac{K (\Delta T)^3}{6} + C_2 + C_3 x
\end{align*} \]

where

- \( M \) = tappet/cam reciprocating mass
- \( \Delta T \) = duration of impacting force
- \( K \) = rate of rise of tappet sideways force
- \( F \) = maximum tappet sideways force
- \( C_1 \) and \( C_2 \) are constants

For tappet to travel clearance \((S)\) to bore:

\[ M s = \frac{K (\Delta T)^3}{6} \]

\[ \Delta T = \left( \frac{6 M S}{K} \right)^{1/3} \]

\[ V_i = \frac{K (\Delta T)^2}{2M} \]

\[ F_i = \frac{2M V_i}{\Delta T} \]

where

- \( F_i \) = tappet impact force
- \( V_i \) = tappet impact velocity

F.F.T. of \( F_i \) is included in this upgraded program (GORIAL-HADDAD 1985) and it was used to predict the effect of realistic tappet/bore wear range, optimum tappet/bore dimensions, and other tappet/bore/cam parameters.

CONCLUSION

Analytical and experimental studies of In-line fuel pump tappet slap have shown that In-line F.I.E. cavitation noise and vibration can be controlled mainly by adopting optimum tappet/bore clearance and the avoidance of excessive wear.

REFERENCES

NOISE EMISSION FROM ROCK IMPACTS

D. O. Hughes and A. J. Doyle

Physics Department, North Staffordshire Polytechnic, College Road, Stokey-on-Trent, United Kingdom.

INTRODUCTION

It is well known that rock and coal preparation plants are inherently noisy, producing environmental conditions likely to cause noise induced hearing injury to employees. Much has been done to reduce noise levels associated with screens, crushers, and conveyors using proven acoustic treatment methods.

There is, however, a need to study the various mechanisms associated with the processes of rock preparation at a fundamental level from the point of view of noise reduction and the design of more efficient processes.

The work described in this paper is a study of one aspect of noise generation in the rock and coal preparation industry, namely noise emission resulting from the impact of rocks. The aim of the project was to identify the major noise producing mechanisms during rock impacts and to develop a theoretical model to describe the resulting sound pressure-time profiles. In order to simplify the problem initially, the project was confined to studying the impact of rocks on thick steel plates.

THEORETICAL CONSIDERATIONS

Mechanisms of Noise Generation

There are five basic types of possible sound generation mechanisms associated with rock impacts.

Aerodynamic noise is mainly due to rapid air movement at impacting surfaces and requires large contact surfaces for it to be observed. The ringing of structures resulting from impacts is almost always observed, but the frequencies of the noise generated by the ringing within the rock samples are normally above the audio range. Collision processes on a screen or conveyor are unlikely to induce fracture noise, but could be a contributing factor in crushing machines. Surface acceleration, caused by impacting surfaces, may give rise to pressure waves important in the case of large surfaces but less so for rock particles. Finally rigid body radiation is a disturbance resulting from the rapid acceleration of the areas of contact during collision. Of these five mechanisms, rigid body radiation is the most important initial noise source in rock-rock and rock-metal impacts.

Impact of Rocks on Thick Plates

When a rock strikes a surface, the accelerations of the contact surfaces are typically 30,000 m/s². This acceleration is associated with a high localised strain energy. The theory for predicting the type and magnitude of the localised deformation was originally developed by Hertz(1). The generalised Hertzian force-deformation law is

\[ F = k_a a^{3/2} \]  

(1.)

Where F is the force acting between the impacting surfaces and a is the distance between the centres of the ellipsoids of contact, and \( k_a \) is dependent on the contact profile and the elastic constants of the impacting materials.

The impact of rocks of random shape on a thick steel slab can be approximated to either point (spherical) or edge (line) impacts. Accordingly, the impact of (a) spherical rock samples and (b) cylindrical rock samples on thick plates were considered. In order to calculate the acoustic pressure due to elastic deformation resulting from an impact, it is necessary first to evaluate the maximum force, maximum acceleration and contact time associated with the deformation.

In the case of spherical rocks, following the work of Koss (2) for metal spheres, the maximum force \( P_m \) and acceleration, \( a_m \), are given by

\[ P_m = \frac{k_1}{2} \left( 1.25 \frac{V_o^2}{L} \right)^{1/4} / (4 k_1 k_2) \]  

(2)

\[ a_m = P_m / M \]  

(3)

Where \( k_1 \) is a constant dependent on the masses, \( V_o \) is the maximum velocity prior to impact and \( M \) the impacting mass.

A line impact may be generated by a cylindrical impact on a flat plate. In this case the Hertzian equation becomes

\[ a = \frac{6}{\pi} \ln \left( \frac{\pi R}{L} \right) \]  

(4)

Where \( R \) is a constant dependent on the elastic constants, \( P \) is the force per unit length and \( L \) is the radius of the cylinder. In order to arrive at expressions for the maximum acceleration, \( a \), force \( P \) and impact time \( T \), iterative methods were used to evaluate appropriate values for the logarithmic term in equation 4. The resulting expressions for \( a_m \) and \( T \) are,

\[ a_m = (V_o \ln(\pi R/L)); \quad T = (\pi^2 \ln(\pi R/L))^{1/2} \]  

(5)

Where \( m \) and L are the mass and length of the cylinder, and \( V_o \) is the appropriate value for the logarithmic term.

Sound Pressure

Kirchhoff (3) originally derived an expression for the acoustic pressure at a point due to the acceleration of a surface. Two equations applying to point sources were used to predict the pressure at a point \((r, \theta)\), one equation applying for a time period during impact and the other for the time period after impact. In the case of the impact of rocks on a flat steel plate dipole image sources need to be considered (Kàyà 4). The resulting equations for the pressure at a point \((r, \theta)\) are lengthy and for this reason are not included in this paper.

EXPERIMENTAL

Sound pressure measurements were taken of the noise generated when rocks of various shapes and sizes were allowed to fall with known velocity onto a steel plate 60x60x2 cm in dimension. The arrangement, together with the measuring instrumentation which included an Analogic Data 6000 analyser, was set up inside an anechoic chamber. Spherical, cylindrical, and cubical steel and rock samples were cut and machined from mild steel, coal, shale, sandstone, granite and limestone. The spheres were between 2 and 4 cm radius and the cylinder radii varied from 1.85 to 2.5 cm, and cylinder length from 1 to 4.5 cm. Values of elastic constants were experimentally determined or obtained from the literature. A number of suspension methods were tried out and the cantilever method found to give the most consistent results. The contact time was determined using a miniature accelerometer attached to the rock sample.
DISCUSSION OF RESULTS

Theoretical values of the acoustic pressure were calculated for the steel and rock samples used. Pressure-time profiles showed three distinct peaks and pulse durations of the order of 1 ms. The effect of gross errors in the elastic constants of a particular rock sample produced only a small (3 dB max) variation in the value of the pressure in the theoretical model.

Experimental determination of acceleration profiles showed curves having a single peak and pulse durations of the order of 0.3 ms.

The experimentally measured pressure profiles showed three distinct peaks similar to those predicted by theory. Figure 1 shows typical acceleration curves obtained for both spheres and cylinders. Figure 2 shows a typical double impact acoustic pressure profile. The values of the peaks of the pressure waves are of the order of 0.4 ms. Integration of the pressure wave gives an energy equivalent of 120 dB for the pressure pulse.

The following is a comparison of data for a 3 cm radius coal sphere with an impact velocity of 0.99 ms⁻¹.

<table>
<thead>
<tr>
<th>Experimental (dB)</th>
<th>Theoretical (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Peak</td>
<td>118.2</td>
</tr>
<tr>
<td>2nd Peak</td>
<td>122.3</td>
</tr>
<tr>
<td>3rd Peak</td>
<td>112.8</td>
</tr>
<tr>
<td>Duration</td>
<td>722μs</td>
</tr>
<tr>
<td></td>
<td>852μs</td>
</tr>
</tbody>
</table>

The following table gives a comparison of the predominant (2nd) pressure profile peak for a number of 2 cm radius spheres having impact velocities of 1.4 ms⁻¹.

<table>
<thead>
<tr>
<th>Sphere</th>
<th>Experimental (dB)</th>
<th>Theoretical (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steel</td>
<td>129.8</td>
<td>128.8</td>
</tr>
<tr>
<td>Sandstone</td>
<td>129.3</td>
<td>128.1</td>
</tr>
<tr>
<td>Granite</td>
<td>126.7</td>
<td>128.0</td>
</tr>
<tr>
<td>Coal</td>
<td>124.1</td>
<td>123.6</td>
</tr>
<tr>
<td>Shale</td>
<td>129.6</td>
<td>129.1</td>
</tr>
<tr>
<td>Limestone</td>
<td>132.8</td>
<td>130.7</td>
</tr>
</tbody>
</table>

Typical comparison of the second peak pressure for cylinders having a radius of 2.5 cm and length 4 cm with impact velocities of 0.99 ms⁻¹ are:

<table>
<thead>
<tr>
<th>Cylinder</th>
<th>Experimental (dB)</th>
<th>Theoretical (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sandstone</td>
<td>126.4</td>
<td>128.8</td>
</tr>
<tr>
<td>Granite</td>
<td>125.1</td>
<td>129.5</td>
</tr>
<tr>
<td>Coal</td>
<td>123.7</td>
<td>126.0</td>
</tr>
<tr>
<td>Limestone</td>
<td>124.4</td>
<td>130.5</td>
</tr>
</tbody>
</table>

As expected the acoustic pressure falls with decreasing impact velocity, e.g. 8 dB per halving of impact velocity.

Comparison of the pressure profiles of spheres and cylinders having similar impact energies show some similarity as for example in the case of a sandstone sphere and two cylinders of sandstone of length 4 cm and 2 cm respectively.

<table>
<thead>
<tr>
<th></th>
<th>4cm Cyl.</th>
<th>2cm Cyl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Peak</td>
<td>120.5</td>
<td>123.6</td>
</tr>
<tr>
<td>2nd Peak</td>
<td>125.6</td>
<td>126.4</td>
</tr>
<tr>
<td>3rd Peak</td>
<td>117.0</td>
<td>120.9</td>
</tr>
<tr>
<td>Duration</td>
<td>419μs</td>
<td>345μs</td>
</tr>
</tbody>
</table>

As expected the acoustic pressure falls with decreasing impact velocity, e.g. 8 dB per halving of impact velocity.

Experimental evidence shows that for spherical impacts, radiation is hemispherical at about 15 cm distance from the contact point. Accordingly, the sound power level of the source lies in the range 106 to 112 dB. In the case of cylindrical impacts there is a small variation of acoustic pressure as the microphone position is rotated about the cylinder position. The pressure decreases as the angle with the normal to the impact plate is increased and when measurements are made along the cylinder axis. Rigid body radiation noise has a typical continuous spectrum with most of the pressure energy appearing in the range 2 kHz to 6 kHz.

CONCLUSION

When rock impacts on a surface, the contact geometry may be regarded as either "point" or "line". Random rock impact may involve multiple point and/or line impacts. The primary source of noise resulting from impact is due to surface deformation and is identified as rigid body radiation. The most important factor governing the magnitude of rigid body radiation pressure is the impact velocity of the rock. The theoretical treatment of rigid body radiation has yielded a model which gives results that are in close agreement with experimental values for both spherical and cylindrical rocks impacting on a thick, rigid slab of material.

REFERENCES

GERÄUSCHMINDERUNG TIEFFREQUENTER IMMISSIONSANTEILE IM HAUSHALTS- UND FERNBEREICH VON NASSIRIBEN

H.-U. Haering, J. Schmitz

BFI Betriebstechnik GmbH, Sohnstraße 65, 4000 Düsseldorf 1


In einem ca. 660 m entfernten Wohngebiet wurde über die Vibration der Fensterscheiben geklagt. Aus Bild 1 ist die Lage des Betriebes und der Wohnbebauung zu ersehen. Die Bewohner waren der Ansicht, daß die Vibration der Fensterscheiben durch Bodenscherungen verursacht werden. Diese Vermutung konnte jedoch durch die Messung der Schwingschwingaktivität an den Wohnhäusern mit Notfilter schwinggeschwindigkeitssenden von 95 widerlegt werden.

Im Auftrag des Betreibers der Siebe wurden daher im Umfeld der Anlage und in der Wohnbebauung Luftschallmessungen durchgeführt mit dem Ziel, neben der Ursachenerkennung auch Verminderungsmaßnahmen vorschlagen. Es wurde geschehen, daß neben anderen Maßnahmen die Geräusche der Schwingfördersäule von den Mitarbeitern subjektiv als störend angesehen wurden.

Die Anordnung der beiden Siebe innerhalb des Werkes ist aus Bild 2 ersichtlich. In den Bildern 3, 4 und 5 ist die Konstruktion mit den beiden Antrieben sowie die Aufstellung gezeigt.

Messungen des Luftschallpegels ergaben keinen Aufschluß über die Ursache. Daher wurde an einem Meßpunkt an der Werksgrenze (Bürogebäude) bei Ausmung einiger Quellen eine Frequenzanalyse des vorliegenden Geräusches durchgeführt.

Die spektrale Verteilung für den Bereich von 5 Hz bis 5 kHz am Meßpunkt Bürogelände ist in Bild 6 dargestellt. Bei der Messung war auffallend, daß die beiden Tretzfläche von 16 und 20 Hz aus dem Spektrum herausragten und in ihren Amplituden bis zu 25 dB (etwa 70 bis 95 dB) schwankten.


Da diese Reduzierung zwar für die Wohnbebauung ausreichend war, jedoch im Werksbereich noch immer zu Beschwerden führte, mußten weitere Maßnahmen untersucht werden. Es wurde vorgeschlagen, die beiden getrennten Antriebe durch einen zu ersetzen und durch eine Übergangsmassen mit einem vorgestellt, daß die Schwungung von 12 Hz sowie die große Amplitude nicht mehr vorhanden ist, was subjektiv zu einer sehr erheblichen Lärmentlastung beiträgt.
LEVEL OF BUS NOISE AND INFLUENCES OF ENGINE SUCTION AND EXHAUST SYSTEM

V. Tandara

Dr.-Ing. V. Tandara: Technical consulting, development and designing; Expert for combustion machines and reduction noise of vehicles. D-1000 Berlin 10, Postfach 1323

Introduction

The traffic noise is a problem to which a particular attention is paid in towns. A great portion of the traffic noise is the result of the use of buses which are increasingly used in the town traffic. The most important reasons for the increased use of buses in the town traffic are as follows: decrease of the traffic in the town center, reduction of emission of exhaust gases and, consequently, reduction of noxious components in them, limited number of parking places etc.

The figure 1 gives the previous, present and future limit values for various types of vehicles. What strikes the eye is the fact that there will be a difference of 5 dB(A) only between the limit values for the passenger cars and the heavy utility vehicles which include also certain types of buses. The principal reason for that discrepancy should be sought particularly in the lack of awareness of the influence of the individual sources of noise on the total noise of the vehicle /1, 2/.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PASSENGER CARS</td>
<td>82</td>
<td>80</td>
<td>75</td>
</tr>
<tr>
<td>BUSES TO 3.5 t</td>
<td>84</td>
<td>81</td>
<td>76</td>
</tr>
<tr>
<td>BUSES ABOVE 3.5 t</td>
<td>89</td>
<td>82</td>
<td>80</td>
</tr>
<tr>
<td>BUSES ABOVE 147 kW</td>
<td>91</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>UV TO 3.5 t</td>
<td>84</td>
<td>81</td>
<td>78</td>
</tr>
<tr>
<td>UV ABOVE 3.5 t</td>
<td>89</td>
<td>86</td>
<td>80</td>
</tr>
<tr>
<td>UV ABOVE 147 kW</td>
<td>91</td>
<td>88</td>
<td>80</td>
</tr>
</tbody>
</table>

Fig. 1: Limits of external noise of the vehicles in dB(A)

Types of buses and sources of noise

The buses are made for various purposes, they are of various sizes and have different engine outputs. They are the town, suburban and tourist buses of a total length of 6, 8, 10, 12 and more than 12m. The most important parts of any bus are as follows: driving unit, transmission, chassis, superstructure etc. The driving unit can be installed in the front, in the middle of the bus or in the rear. The individual internal and external engine noise sources were dealt with in articles /3, 4/.

Location of the driving unit has a significant influence on the bus noise level, particularly on the internal noise. If the driving engine is located in the front with the drive on the rear axle the transmission are incorporated along the entire bus. The influence of the transmission and, also, the influence of the driving engine on the noise level is strong in all bus parts. This applies also for the influence of the engine suction system which is located in the bus front, too, figure 2. Depending on the bus length the exhaust system terminates in the second third of the bus or in the bus rear part.

Fig. 2: Chassis of the bus A9 /5/

Engine suction system

The engine suction system is an important component part of any combustion engine. It supplies air for fuel combustion to the engine operating cylinders. The influence of the suction system on the engine noise level and thus on the bus noise level can be considerable and depends much on the place of incorporation in the bus. In addition to the row engines having only one branch of the air supply line it is necessary to mention the V-engines requiring two lines, figure 3. The suction system of any combustion engine consists of one or two suction collectors, one or more air filters and required piping. On the buses with considerably reduced noise level the sound silencers are additionally incorporated.

One V-engine located in the rear of the bus requires two air supply lines. They are located in the rear on both bus sides. The openings of the air supply lines for combustion are located in the channels for air supply for the engine cooling or above these
channels at the level of heads of the passenger sitting in the last row seats in the bus. The location of the opening for the air supply for combustion depends in particular on the surroundings in which the buses will be used, i.e., on the quantity of dust which is to be expected during their use. Because of the impulse streaming in the suction system resulting from the periodic operation of the valves the resonances may occur on the bus body limit surfaces and may condition an increase of the bus noise. The low frequency vibrations of the air column in the air supply line occurring during the engine empty-load run are the most inconvenient and that not only because the influence on the people but also because of the limited possibilities of the effective reduction of such low frequency noises.

The figure 4 gives the internal noise level of the bus A11 depending on the engine number of revolutions on which the external noise level is reduced to approx. 50 dB(A). The noise level is measured on the last row seats i.e., on the outside left and middle seats. Although the max. noise level of 77-78 dB(A) was very low there was a certain difference between the noise levels on seats. The principal reason of those deviations should be attributed to the influence of the suction system whose supply openings were located at the level of the head of the passengers.

The figure 5 gives the external noise level round the bus A11P standing still before and after reduction of noise. The noise is measured at maximum engine speed of 2650/min at 7m distance. In the area of the suction system opening and the engine cooling air supply opening the values measured are 79 dB(A). These values do not differ from the values measured on the other places round the bus.

Engine exhaust system

The combustion engine exhaust system is consists of one or two exhaust collectors, one or more sound silencers and required piping. Its influence on the engine and bus noise level was very important as long as the permissible external noise level of heavy buses was 89-91 dB(A). In that period it was considered to be the most important noise source not only of the engine but also of the bus. That was reflected not only in the determination of the future limit values of the vehicle external noise but also in working out of the rules and regulations for measuring of external noise according to ISO 1120/1,2.

The bus A8 with the engine incorporated in the front where after reduction of the external noise to 80-81 dB(A) in the area of the exhaust system a considerably lower noise level was measured than in the bus front part, figure 6/5, best shows that the combustion engine exhaust system is only one of many engine noise sources. The noise level is measured at 7m distance round the bus at the max. engine speed of 2950/min. Something similar can be seen in figure 5 where the noise level round the bus A11P rear part is approximately the same on all measuring points.

Conclusion

The suction and exhaust systems are to out of many combustion engine and bus noise sources. On the buses with the external noise level reduced to 80 dB(A) it is not possible to expect a considerable reduction of the vehicle noise level only by working on the engine suction and exhaust system but it is necessary in the same time to reduce the influence of other engine and bus noise sources.

References

INTRODUCTION

Textile spinning operations are known to be quite noisy with noise levels ranging from 90 to 100 dB in most ring spinning rooms. The primary noise sources have been found to be associated with the spindle-bobbin system, the ring-traveler system, and the vacuum-end collection system. Several experimental studies have been undertaken to reduce the contributions of the ring-traveler system and the vacuum-end collection unit into the overall noise. On the other hand, the spindle-bobbin system remains as the least investigated unit with a predicted possible noise reduction up to 10 dB [1]. The sound radiation characteristics of the spindle and bobbin needs to be studied in detail before design changes on the various spindle components have been proposed. Early studies have been directed to determine the effect of the eccentricity of the spindle-bobbin system rotating at high speeds. The most popular model has been a rotating dipole developed by Crawford [2]. The expressions for the intensity and the sound power have been obtained and discrete frequency noise of an eccentrically rotating bobbin has been determined using this model. It has been demonstrated later by an experimental study that the noise is directly proportional to the clearance between the spindle and bobbin when the eccentricity used in the dipole model is taken as the value of this clearance at the base of the bobbin [3].

The noise produced by the rotating yarn balloon is also of great concern. A compact source model composed of a rotating force distribution has been proposed by Handschy [4]. Sound radiation of a yarn balloon represented by a single dipole set at the center of the spindle of a two-for-one twist-winder has been studied following the theoretical acoustics of a rotating point force. The same problem has been dealt [5] by applying the general expression for the sound field of a point force in arbitrary motion given by Lowson [6] to the rotating yarn balloon of the same type. After the balloon shape has been determined, the sound pressure at an observation point has been obtained by assigning a point source to each straight yarn element and summing up the individual contributions of each source. The frequency domain information has also been made available by taking the FFT of the resulting pressure-time history [5].

In this study, the contribution of a rotating spindle-bobbin system with a linearly varying mass unbalance into the overall noise is determined. The forced vibrational behavior of the system is obtained by treating it as a rotor-bearing system and by using finite elements [7]. The calculated whirl shape is used to determine the sound radiation characteristics by the aforementioned approaches on balloon noise. Two different models are developed to evaluate the contribution of the noise radiated by the spindle-bobbin system moving in circular whirl orbits at the first forward whirling speed. 

VIBRATIONAL BEHAVIOR OF SPINDLES

The Computer Program

Any rotor system can be represented with a combination of three types of elements. These elements are rotors with axial symmetry and uniformly distributed inertial and elastic properties, discrete bearings, and rigid symmetric disks.

The computer program ROTVIB [8] includes the effects of rotary inertia, gyroscopic moments, axial load, internal viscous and hysteretic damping, and transverse shear deformations in the same model. The data input consists of the geometric and material properties of the system subelements, the spin speed of the rotor, and the accuracy and control parameters. The required number of whirling speeds (forward and backward) with the corresponding logarithmic decrements and modeshapes, and the whirl orbits at the required nodes are the output of the program.

Modeling

Main components of a ring spinning spindle are shown in Fig. 1. The rotating part of the spindle is divided into 13 finite rotor elements made of steel and aluminum as shown in Fig. 2. The whorl is taken as a rigid disk. The footprint bearing is taken as almost rigid by assuming high stiffness values for the specific spindle at hand. The stiffness coefficients of the roller bearing which is located at the top of the rotor element k4 are calculated [7].

The spindle is assumed to be perfectly concentric. The bobbin is divided into six parts and each part is modeled as a rigid disk having an eccentricity. The mass of the bobbin is taken as 0.1 kg, and a linearly increasing eccentricity of the bobbin is defined in such a way that it is 1 mm at the base and 2 mm at the top. The unbalance response of the spindle-bobbin system due to this eccentricity is calculated at all nodes above the roller bearing by using ROTVIB. All the whirl orbits are circular since the isotropic bearings are specified. The shape of the whirl is determined by using the least squares method.

MODELS FOR SOUND RADIATION

The compact source models for sound radiation assume that the spindle-bobbin system follows the shape of its first whirling mode at a speed equal to its first forward whirl speed with zero spin speed. In other words, the noise contribution due to spin is excluded in the study.

The radiated sound power of the nth harmonic \( W_n \) and its corresponding directivity \( Q_n \) for free field conditions are given by Handschy [4] as

\[
W_n = \frac{\rho c \omega^2 C_n}{2} \frac{m}{C_n} (n+1)^4 a_n
\]

where

\[
\sigma_n = \frac{1}{2\pi} \left( \frac{n+1}{2} \right) \left( \frac{2n+2}{2} \right) \sin^{2n+1} \theta \left| I_n \right|^2 d\theta
\]

\[
I_n = \int_0^\infty C_0 \sqrt{1 + \left( \frac{dr}{dz} \right)^2} e^{-in\xi \cos\theta/c} dE
\]

\[
\omega = \frac{1}{k^2} \sum_{n=0}^{\infty} \frac{(n+r)sin\theta}{2c} \left( \frac{dr}{dz} \right)^2 dE
\]

\[
\sum_{k=0}^{\infty} \frac{1}{k!(n+k)!} dE
\]
\[ Q_n = \frac{1}{\rho_0 V_n^2} \left( \frac{1}{2} \pi d^2 \frac{n^2}{2\pi} \right) \left( \frac{1}{8} I_L \right)^2, \]  

and \( \rho_0 \) is the average air density, \( c \) is the speed of sound in air, \( d \) is the average diameter of the spindle-bobbin system, \( r_n \) is the maximum whirl radius, \( C_p \) is the drag coefficient, \( \omega \) is the first forward whirl speed, and \( \tau_n \) is the coordinate along the axis of the spindle.

In the second model, basic acoustical equations of the sound field radiated by a time varying point force in arbitrary motion derived by Lowson [6] are used. The far field sound pressure \( p \) from a point source in arbitrary motion is given by

\[ p = \frac{1}{4\pi R^2 c (1 - \rho_\nu^2)} \left( \frac{\tau}{R} + \frac{M_2}{1 - \rho_\nu^2} \right) \]  

where \( \vec{R} \) is the vector from the position of source to the observer, \( \vec{F} \) is the Mach number vector, \( M_2 \) is the component of acceleration of the point force in the direction of the observer, and \( \vec{F} \) is the aerodynamic point force. The aerodynamic force has a constant drag component \( \vec{D}_{dc} \), and fluctuating lift \( \vec{L} \) and drag \( \vec{D}_c \) components. Neglecting fluctuating components, one gets

\[ \vec{F} = \vec{D}_c = \frac{1}{2} \rho_0 C_D \omega^2 \lambda^2 \]  

where \( \lambda \) is the length of the rotor element.

RESULTS AND CONCLUSIONS

The finite element model shown in Fig. 2 is employed to obtain the vibrational behavior of the spindle with linearly varying mass unbalance due to the bobbin mounted along its axis. Damping due to both bearings and the oil contained in the bolster is neglected in the analysis. The forward whirling speeds and whirl orbits are calculated for a typical spin speed of 12000 rpm and are calculated by the computer program ROTVIB on a Burroughs B6900 computer at the Computing Center of Middle East Technical University. The first forward whirling speed is found to be 2690 rpm. The equation of the whirl shape due to unbalance forcing is determined by the least squares curve-fitting technique as

\[ r_c(z_c) = 0.0031 \lambda_c^{1.0086} \]  

The first compact source model for sound radiation which uses the equations given by Handschey [4] yields a sound power of \( 1 \times 10^{-11} \) W for the first harmonic. On the other hand, a sound power value equal to \( 2.3 \times 10^{-14} \) W is obtained from the second model which employs Lowson’s approach [6]. The reasons for the very low sound power values are that the whirling speed is relatively low and the whirl orbit radii are very small. The difference in the calculated sound power values is due to the fact that the first model would set an upper limit for the radiated sound power as demonstrated in [4] and the second model neglects the fluctuating force components on the rotor elements moving in whirl orbits.

The directivity patterns are almost the same as illustrated in Fig. 3. Both directivity plots are drawn for one side of the spindle since there exists an angular symmetry about the geometrical axis of the spindle. The resulting directivity patterns possess typical dipole directivity characteristics as expected.

When the rotating dipole model developed by Crawford [2] is applied to the spindle-bobbin system with the same eccentricity and a spin speed of 12000 rpm, sound power generated due to the spinning of the eccentrically rotating bobbin is found to be \( 1.6 \times 10^{-15} \) W. Consequently, it can be stated that the noise generated by the spindle-bobbin system moving in whirl orbits can be neglected when compared to the noise generated by its spin.

REFERENCES


Fig.1. Ring Spinning Spindle  Fig.2. Spindle Model  Fig.3. Directivity Patterns
NOISE REDUCTION OF THE INDUSTRIAL WASHING-MACHINES

P. Mandl; V. Tandara
* TU PHILMAF, ELMNI, YU-62000 Maribor, Industrijska 22
** Dr.-Ing. V. Tandar: Technical consulting, development and designing,
D-1000 Berlin 13, Postfach 1323

Introduction

In the everyday life great quantities of filthy linen are caused. For cleaning that linen mainly the washing machines are used. Household and industrial washing machines are distinguished depending on the purpose and quantity of the filthy linen.

The capacity of the household washing machines is up to max. 10 kg of filthy linen in one washing cycle and they are used particularly in the individual households. They are placed in kitchens, bathrooms or basements depending mainly on the existing possibilities of the individual households.

In industrial washing machines are used for washing large quantities of linen. These are the washing machines in which large quantities of linen can be washed in one washing cycle. The usual capacities of those washing machines are 15, 20, 40, 100 or more than 100 kg. Such machines are set up in the laundries of hospitals, factories and plants occupied with cleaning large quantities of linen of various kinds.

A great number of the same or different washing machines, figure 1 and 2, is usually located in the laundry rooms. In addition to the washing machines also other machines are provided in laundries, such as linen driers or large industrial flat-irons. Great lines where the major part of work is automatized are fitted in modern laundries for continuous washing of large quantities of linen in one or more work shifts during a day.

Each operation of the industrial washing machines is connected to a series of problems which have to be solved. One of the very important problems is the high noise level caused not only by the large industrial washing machines but also by the small household washing machines.

Noise source in laundries

The small household washing machines cause an average noise level of 65-70 dB(A). The noise levels of the industrial washing machines can be even more than 90 dB(A). This value is obtained by measuring at a distance of 7m. In one household a washing machine noise level of 65-70 dB(A) is not low and can be heard in the neighbouring rooms. This is the case particularly if the limit surfaces of the room do not fulfill the legal requirements concerning propagation of noise. That occurs in houses built in a time when such an attention was not paid to the problem of the noise as nowadays.

In the large industrial laundries one room accommodates a great number of washing machines, linen driers and large industrial flat-irons. All these machines cause the noise, however, the highest noise level occurs during operation of the washing machines. That washing machine level can increase if a great number of washing machines is in operation in the same time. As it is well known two noise sources of the same intensity increase the noise level for 3 dB(A) and four noise sources for 6 dB(A). That means that four washing machines of 84 dB(A) noise level each in operation would cause a noise of 90 dB(A) in the room. The actual noise level can be affected also by other factors; the most important factors are: shape of the room, arrangement and position of the individual machines, quality of the limit surfaces of the room etc.

As the industrial laundry walls and floors are lined with ceramic tiles the sound reflection and consequently, the noise level increase in the room will occur. The similar applies in the household kitchens and bathrooms therefore the actual noise level in them will be considerably higher than the noise level of the machine itself. Nevertheless, the noise level itself will be smaller than the noise level in laundries. While in the individual dwelling houses a small number of people is exposed to the washing machine noise level, a considerably greater number of people is exposed to it in industrial laundries. On the individual places the noise level can exceed the permissible limit value of 85 dB(A) thus the health of the worker concerned is
endangered. The table 1 gives the permissible noise levels on the work places.

Washing machines

The design concept of all washing machines is similar. They consist of a supporting structure and a housing. Depending on size the machines are put on rubber or metallic springs. The principal parts of the washing machines are: rotating drum, electric motor, electric motor support, coupling, belt pulleys, heating and an electronic device for control of the process. Depending on the purpose and size of the machine one or several electric motors are incorporated. They are of different outputs and speeds. The electric motors are placed on special supports and serve for rotating the drum. The drum rotates via the belt pulley, belts and coupling. After the drum has been filled with filthy linen, water and detergent, heating is switch on and the drum is put into operation. The drum rotation speed is different; it depends on the type of washing. Usually, the following parts of the process are distinguished: sorting of linen, washing of linen, low and high centrifugal rotation.

The washing and washing are effected at low speed. The driving electric motors are of low power.

The low and high centrifugal motion is effected at a considerably higher speed so that for these parts of the process considerably stronger electric motors are necessary. The electric motors can be of one or multiple speed type, i.e., 2, 4, 8-pole or more. On the small industrial washing machines it is possible to use only one multiple-speed motor. On the large washing machines the electric motors of different outputs and speeds are required.

Washing machine noise sources

The production programmes of the PRJAT include the K-50, K-70, K-20, K-40, and K-100 washing machines. Their washing capacities are 15, 20, 40 and 100 kg of filthy linen in one working cycle.

The industrial washing machines are of large dimensions; for their operation the electric motors of different outputs are. All this causes that the washing machine noise level is high. The principal washing machine noise sources are: driving units, couplings, superstructure, drive parts, vibrations etc.

The driving units are the electric motors of different outputs and speeds. Their noise level depends not only on the output and the speed but particularly on the design and quality of the manufacture. This can be best seen in the figure 3 indicating the results of the frequency analysis of several motors of the same type. These electric motors are incorporated in the K-20 washing machines and serve for distribution and washing. The frequency analysis is performed in front of the machine at a distance of 1m. The noise level of these electric motors is given in table 2.

The figure 4 gives the results of the frequency analysis of the K-20 washing machine during high centrifugal motion. Both curves indicate the condition of the washing machine before and after the change of the electric motor. For comparison the result of the frequency analysis of the electric motor are stated. The washing machine noise level was approx. 60 dB(A) and the T-112 motor approx. 75 dB(A).

The washing machine noise level during various washing processes is given in table 3. During the start of the low centrifugal motion and the stage of transition from the low to high centrifugal motion the noise level increases. That noise level is higher than the washing machine noise level during constant machine speed. This is caused due to the design of the drive.

The figure 5 gives the results of the frequency analysis of the driving electric motors of a K-50-15 washing machine. As it can be seen, their noise level is considerably lower than the noise level of the electric motors of the K-20 washing machine which affects also the washing machine noise level of 65-70 dB(A). The effect of this noise level can be best seen from the fact that the noise level of a modern washing machine of 4,5 kg capacity is given. Because of the limited extent of this paper are presented some of particulars.

<table>
<thead>
<tr>
<th>No</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>dB(A)</td>
<td>65</td>
<td>64</td>
<td>71</td>
<td>77</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sorting</th>
<th>Washing</th>
<th>Low centr.</th>
<th>High centr.</th>
</tr>
</thead>
<tbody>
<tr>
<td>73/74</td>
<td>61/65</td>
<td>77/79</td>
<td>80/83</td>
</tr>
</tbody>
</table>

Fig. 5: Noise of electric motors T-112 MA/12

Fig. 4: Noise of washing machine PC-20 and motor

Table 2: Noise of el. motors T-112 MA/12

Fig. 5: Noise of el. motors T-112 MA/12

Table 3: Noise of washing machine PC-20 in dB(A)

Fig. 6: Noise of washing machine PC-20 in dB(A)
LONG-TERM MEASUREMENT OF AUDIBLE NOISE AT THE
SHIOBARA HVDC TEST LINE

Y. Nakano and M. Fukushima

Central Research Institute of Electric Power Industry
1-1, Iwato Kita 2-Cho, Komae-Shi, Tokyo, 201 Japan

INTRODUCTION

Over long distances, high-voltage direct-current (HVDC) transmission can wheel large quantities of power more economically than AC transmission. Several DC overhead lines are being operated around the world and some new lines have been proposed or are under construction. Preliminary designs had already been completed in Japan to supply bulk power (about 100kW) at a distance of several hundred km through HVDC lines.

The Shiobara HVDC test line was constructed to the further study to investigate in detail the environmental factors in the vicinity of DC lines. Long-term measurement of corona and field effects such as audible noise (AN), television and radio interference, corona loss, ion current density and electric field at the ground level, induced voltage on insulated objects under the line, and ozone has also begun.

Of these parameters, AN has become recently as an important design consideration. Some institutes have proposed methods for calculating the AN level in the vicinity of DC lines. These methods have been developed with the results of the measurements at test lines.

TEST FACILITIES

Test Line and Power Supply

The Shiobara HVDC test line consists of a 4 x 3.84cm bipolar double circuit with a span of 310m for the measurements using two pairs of movable crossbeams on two towers permit variation of the line height and pole spacings. In the test reported in this paper the line configuration was fixed for a minimum height of lower conductor bundles (N) of 22a, a vertical pole spacing (C) of 16m, a horizontal pole spacing (W) of 22m and a subconductor spacing of 40cm. A polarity arrangement was also fixed in the opposite configuration, given in the insert in Fig.1. For this arrangement the average-maximum bundle gradient factors were 0.0370cm⁻¹ for upper bundles and 0.0382cm⁻¹ for lower bundles.

The power supply consists of two cascade rectifiers to energize the test line in the range from +350kV to +800kV.

Instrumentation and Test Procedures

A five-channel precision sound level meter equipped with five microphones 1.22cm (0.5in.) in diameter and a real-time spectrum analyzer were used for the AN measurement. The microphones were covered with windshields and located at five points beneath the line, as shown in Fig.1. Each microphone was positioned 1.4m above the ground level.

A-weighted AN data and frequency spectra were recorded in the data acquisition system once every minute, along with the date and time, applied voltage, meteorological parameters and other environmental variables. As weather conditions in particular have a considerable effect on AN generation, meteorological parameters including temperature, atmospheric pressure, relative humidity, wind velocity and direction, precipitation and insolation were sampled simultaneously.

Fig.1 Location of microphones (unit: m)

TEST RESULTS

Statistical AN Performance

During the period from September 1982 to September 1983, measurements were carried out intermittently at voltages between +350kV and +500kV under different weather conditions.

In the analysis, AN data that greatly influenced by high ambient noise (twitter of birds, chirp of insects, croak of frogs, artificial noises, etc.) were possibly eliminated. As wind has a great effect on the generation of corona discharge, data taken at wind velocities above 2m/s were also rejected, in order to eliminate wind noise.

Statistical fair weather AN performances of average level and the standard deviation, excluding winter, are shown in Fig.2, where the influence of ambient noise is compensated for. The AN level at voltages below +500kV was almost the same as the ambient noise level of 33dBA.

Fig. 3 shows statistical fair weather octave band frequency spectra, excluding winter. In the figure the typical octave band frequency spectrum of AN from AC transmission lines in the rain (relative value) is also presented for reference. It is a remarkable difference that AC AN has a 100Hz or 120Hz pure tone peak. With the exception of this difference, the two resemble each other.

The cumulative distribution is given in Fig.4. The AN level in winter was several dBA less than in other seasons. There was no meaningful variation in the AN level from spring through summer to autumn.

For AN produced from DC transmission lines, the generation quantity in foul weather is much lower as is shown in Fig.3. According to the figure, the AN level decreased by 10 dBA about 20 minutes after the rain started. This is the most remarkable difference between DC AN and AC AN that increases in the rain.
Fig. 2 Average AN level for increasing line voltage (point AN-3 for spring/summer/autumn, fair weather)

Fig. 3 Octave band frequency spectra (average value at point AN-1 for spring/summer/autumn, fair weather)

Fig. 4 Cumulative distributions of AN (at point AN-3 for fair weather)

Fig. 5 Temporal variation of AN after beginning of rain (at point AN-1)

Comparison between Measured and Calculated AN Levels

Four types of calculation methods for AN prediction in the vicinity of DC lines are summarized in [1]. Two are for average level and the others are for L₁₀₀ and maximum level. The DKIEPI calculation formula has been revised and refined in [2]. As in other methods, DKIEPI's method is based on the general fact that most AN generation occurs on positive conductor bundles and that generation quantities from negative pole can be neglected from a practical point of view. But as these methods are for single circuit lines, their application to double circuit transmission lines requires that the contributions of two positive poles are compounded at the point where the calculation is carried out.

A comparison between the measured and calculated AN levels is made in Table 1. At the higher voltage range to give a rather stable AN generation, the AN levels calculated with the DKIEPI and BPE (for spring and autumn) methods show a good agreement with the measured values. The value calculated with IREQ's method is a little lower as a whole. The maximum value calculated with the PGH method is compared with the measured L₁. The calculated one is generally slightly higher.

It is not an aim to evaluate these calculating methods, because they are based on the results of the measurements at respective test lines, differences must be due to differences in climate, conductor surface conditions, and other factors. It is important to identify them.

Table 1. Comparison between measured and calculated AN levels (at point AN-3, +650kV, fair weather)

<table>
<thead>
<tr>
<th></th>
<th>DKIEPI</th>
<th>DKIEPI</th>
<th>IREQ</th>
<th>EPA</th>
<th>FOR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SPRING/AUTUMN</td>
<td>SPRING/AUTUMN</td>
<td>SPRING/AUTUMN</td>
<td>SPRING/AUTUMN</td>
<td>SPRING/AUTUMN</td>
</tr>
<tr>
<td>L₁₀₀</td>
<td>45</td>
<td>46</td>
<td>42</td>
<td>42</td>
<td>----</td>
</tr>
<tr>
<td>L₁₀₁</td>
<td>46</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>L₁</td>
<td>50</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>52</td>
</tr>
</tbody>
</table>

REFERENCES


2) M. Fukushima, et al., "Prediction Method and Subjective Evaluation of Audible Noise Based on Results at the Shibaura HVDC Test Line", to be presented at IEEE/PES Meeting.
RESEARCH ACTIVITIES ON TRANSMISSION LINE AUDIBLE NOISE IN JAPAN

M. Fukushima, K. Tanabe, Y. Nakano, and S. Tomita

Electrical Engineering Laboratory, Central Research Institute of Electric Power Industry (CRIEPI)
2-11-1, Iwatokita, Komae-shi, Tokyo 201 JAPAN

Studies on the noises from high voltage electric power transmission lines began in Japan at the end of 1960's. Extensive research on audible noise (AN) and acolian noise from conductors and insulators for both AC and DC transmission lines has been conducted by using the test facilities shown in Table 1.

Table I TEST FACILITIES FOR NOISE FROM ELECTRIC POWER TRANSMISSION LINES

<table>
<thead>
<tr>
<th>TEST FACILITY</th>
<th>SPECIFICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akagi UHV test line</td>
<td>max. voltage: AC 1400kV</td>
</tr>
<tr>
<td></td>
<td>length: 600m (2-span)</td>
</tr>
<tr>
<td></td>
<td>2-circuits, 3-phases.</td>
</tr>
<tr>
<td>Takeyama 500kV test line (-1984)</td>
<td>voltage: AC 500kV</td>
</tr>
<tr>
<td></td>
<td>length: 500m (2-span).</td>
</tr>
<tr>
<td>UHV corona cage</td>
<td>length: 24m</td>
</tr>
<tr>
<td></td>
<td>cross-section: 8m x 8m.</td>
</tr>
<tr>
<td>EHV corona cage</td>
<td>length: 55m</td>
</tr>
<tr>
<td></td>
<td>cross-section: 4.2m circle</td>
</tr>
<tr>
<td>Shibara HVDC test line (1982-)</td>
<td>max. voltage: DC ±600kV</td>
</tr>
<tr>
<td></td>
<td>length: 750m (3-span)</td>
</tr>
<tr>
<td></td>
<td>2-circuits with movable arms.</td>
</tr>
<tr>
<td>Shibara HVDC test line (1980)</td>
<td>max. voltage: DC ±500kV</td>
</tr>
<tr>
<td></td>
<td>length: 2410m (4-span)</td>
</tr>
<tr>
<td>UHV fog room</td>
<td>dimension: 35m x 20m x 35m</td>
</tr>
<tr>
<td></td>
<td>max. voltage: AC 900kV (L-G)</td>
</tr>
<tr>
<td></td>
<td>DC ±500kV</td>
</tr>
<tr>
<td></td>
<td>fog generator: 2400kg/h x 2</td>
</tr>
<tr>
<td>EHV fog room</td>
<td>dimension: 20m x 20m x 20m</td>
</tr>
<tr>
<td></td>
<td>max. voltage: AC 600kV (L-G)</td>
</tr>
<tr>
<td></td>
<td>DC ±400kV</td>
</tr>
<tr>
<td></td>
<td>fog generator: 1000kg/h</td>
</tr>
<tr>
<td>Wind tunnel (Hитсchi Cable LTD.)</td>
<td>max. wind speed: 75m/s</td>
</tr>
<tr>
<td></td>
<td>nozzle size: 120cm x 200cm</td>
</tr>
<tr>
<td>Wind tunnel</td>
<td>max. wind speed: 30m</td>
</tr>
<tr>
<td></td>
<td>nozzle size: 40cm x 40cm</td>
</tr>
</tbody>
</table>

All facilities except the larger WIND TUNNEL were equipped in CRIEPI.

AN FROM CONDUCTOR CORONA OF UC LINES

Construction in Japan of an ultra-high-voltage (UHV) transmission line with a maximum voltage of 1100kV will begin by the end of 1980's. The CRIEPI's Akagi 1000kV project on the design of this line started in 1981 [1]. Demonstration tests on corona and field effects at the double-circuit three-phase UHV test line have been completed for conductor bundles of 10 x 8 x 8, and 6 x 3.84cm. The AN long-term test results showed good agreement with predictions.

The predictions were based on data from corona cage tests that were done for a variety of conductor bundles with subconductor numbers of 1 to 10 and diameters of 2.24cm to 5.28cm [2]. Acolian noise from the conductors was also investigated, and the radiation effect of spiral wires wound on conductors was examined. The generation and the propagation of 100Hz hum from the conductor corona is a complicated subject, and present activities are directed towards the study of this phenomena.

AN from 500kV transmission lines, which are in operation all over the country, is not a serious problem because four-conductor bundles have been adopted in order to restrain radio interference strictly. In limited areas, however, the spiral wires required for the reduction of the acolian noise increase AN, especially 100Hz hum. A spiral wire winding method was studied to reduce both AN and acolian noise. A test result of 100Hz hum performance with the corona cage is shown in Fig. 1. The Type II winding shows good characteristics that are almost the same as without spiral wires, whereas Type I increases the hum somewhat. The performance of Type II was also examined at the Akagi test line.

AN FROM CONDUCTOR CORONA OF AC LINES

Studies were made at the Shibara laboratory of CRIEPI in two phases. In the first phase, from 1971 to 1980, short-term tests on the conductor bundles of 1 x 4.94cm and 1 x 4 x 2.53cm were carried out up to +500kV at the old test line. The second phase began in 1982. The long-term tests on 4 x 6 x 3.84cm were conducted up to +800kV at the double-circuit test line. A new method of AN prediction was developed based on the results of the short-term tests. The method is applicable to both bipolar and monopolar lines. The effect of the distance between the positive and negative pole conductors or the pole-spacing is introduced as a parameter. The results of the long-term test are presented in the accompanying paper [3].

AN FROM INSULATORS OF AC LINES

Salt contamination of insulators in transmission lines is not uncommon in the coastal area of Japan. Under high humidity conditions such as in fog or drizzling rain, AN is occasionally radiated from partial discharges on contaminated and wet insulators. The AN performances of various insulator strings of different unit type, unit number in a string, and degree of contamination were investigated at voltages 275, 500, and 1100kV [4]. The tests were mainly conducted in a UHV or EHV fog room. An insulator string with artificial contami-
nation was wetted by artificial fog in the room, and the sound pressure level (SPL) of AN generated from partial discharges was measured. SPL was converted into the acoustical power level (PWL) by a well-known diffused sound field method, as shown in Fig. 2. AN levels around a tower of transmission lines were calculated from the PWL on the assumption that insulator strings generated the same acoustical powers in spherical waves.

Some tests on artificially contaminated insulator strings were made at an outdoor test site under high humidity weather conditions during summer nights at a voltage of 500kV. Field measurements under the tower of the Takeyama 500kV test line were also performed when its insulators were naturally salt contaminated and wetted. SPLs or converted PWLs of AN obtained from the different test methods were nearly the same.

![Fig. 2 SPL measured in the UHV fog room and converted PWL of AN from a contaminated insulator string at 1100/3kV. (number of units 56, salt deposit density 0.1mg/cm²)](image)

AN FROM INSULATORS OF DC LINES

After the Shihara test line was energized in 1871, unusually heavy discharge activities on the DC insulator strings were observed during certain weather conditions such as dense fog and drizzle. The phenomenon, designated as "Single Unit Flashover (SUF)" completely bridges one or two insulator units in a string, and generates extremely strong impulsive sounds. A series of tests was carried out with artificially contaminated insulator strings at an outdoor test area of the Shihara laboratory under natural humid conditions [3].

Since the phenomenon is caused by extreme voltage concentration on units resulting from the thermal balance on the insulator surface, determined by the rate of wetting and drying, the mechanism was investigated from the performance of the insulator unit. Fig. 3 shows the relation of a leakage current through an artificially contaminated insulator against an applied voltage, measured under natural or artificial wet conditions. Under high humidity, at a certain voltage range the leakage current decreased as the applied voltage increased. The unit showed "resistive, negative resistive or discharge" mode as the applied voltage rose, as shown in Fig. 4.

This characteristic was well explained by the following equation:

\[
\frac{dW}{dt} = V_0 - V \times I
\]

Where the left term is the rate of increase in moisture W on the surface of the insulator, W_0 is the moisture supplied from an extremely humid atmosphere and V \times I is evaporated by Joule's heat from the leakage current. The transition of a voltage distribution in an insulator string and break-out process of the SUF when the contaminated surface was wetted were simulated with a computer model based on the above equation.

![Fig. 3 Relation of a leakage current through an artificially contaminated insulator against an applied voltage under wet conditions. (insulator type 250mm suspension, salt deposit density 0.04mg/cm²)](image)

![Fig. 4 Conceptual explanation of characteristic in applied voltage versus leakage current of an insulator unit](image)

REFERENCES


BRUIT EMIS PAR UN OBSTACLE CYLINDRIQUE, PROFIL OU NON, PLACE DANS UN ÉCOULEMENT.

H. Arbe

Institut National de Recherche et de Sécurité
B.P. 27, 54501 Vandoeuvre France

1 - Introduction
Le bruit rayonné par les câbles d'une ligne électrique haute tension devient gênant lorsque la vitesse du vent, soufflant perpendiculairement à la ligne, atteint 16 m/s [1]. Pour prédire le niveau de bruit ambiant, il est nécessaire de connaître avec précision le champ de pression se développant sur cet obstacle [2]. Cette information est difficile à obtenir car les mécanismes impliqués dépendent de nombreux facteurs. Deux d'entre eux jouent un rôle très important dans le cas de lignes électriques : la géométrie des câbles utilisés (très complexe [3]) et la nature de l'écoulement incident (le câble est placé dans la couche limite atmosphérique). On connaît aisément la difficulté à modéliser un tel problème. Cependant, dans le cas d'obstacles profilés, nous disposons de modèles fiables [4], [5], [6], [11]. Il est donc important de tester la transposition au cas d'obstacles non profilés.

Dans ce but, nous considérerons deux obstacles différents : un profil d'aube et un cylindre circulaire. Pour chacun d'eux, nous analyserons les mécanismes responsables du champ de pression induit. Puis nous préciserons l'effet d'une perturbation sur ces mécanismes ainsi que ses conséquences sur le bruit rayonné.

2 - Cas d'un obstacle profilé

a) Écoulement sain
Le spectre de bruit rayonné par un profil d'aube placé en écoulement sain (donc le nombre de Reynolds basé sur le diamètre D est compris entre Re = 60 et Re = 2 000), correspond essentiellement aux composantes [4]. Une composante large bande centrée autour de la fréquence de Strouhal f_s (fig. 1), sur laquelle se superpose une contribution à fréquences discrètes, notée f_c. Le mécanisme associé est le suivant : les instabilités de Tollmien Schlichting du champ de pression induit, puisse des perturbations sur le profil, les ondes sonores ainsi produites se propagent en champ lointain et sont responsables de la contribution large bande observée (f_c). Elles sont aussi à l'origine d'un bouillage aéroacoustique exposé dans [4] qui produit l'émission sonore à fréquences discrètes multiples f_s.

La mesure du champ de pression se développant sur l'aube, confirme cette hypothèse et montre la prépondérance du bord de fuite dans ce mécanisme d'émission sonore [5].

b) Écoulement incident turbulent
Lorsque le taux de turbulence de l'écoulement incident augmente (au delà de 2.5 %), le pic de Strouhal f_s s'estompe et l'on observe alors une augmentation importante du niveau sonore rayonné dans la plage des basses fréquences (située entre 100 Hz et 1 kHz, centrée autour de 200 Hz, pour un profil de 8 cm de cordée et une vitesse d'écoulement de 20 m/s, comme le montre la figure 1). Pour un taux de turbulence voisin de 12 %, la contribution basse fréquence augmente encore et cette augmentation atteint une valeur voisine de 20 dB pour la fréquence dominante de 200 Hz (fig. 1).

Le mécanisme associé à cette émission sonore peut être schématisé de la façon suivante : les fluctuations de vitesse de l'écoulement incident induisent une portance fluctuante sur le profil. Cette portance est alors responsable de l'émission sonore observée en champ lointain.

Une fois encore, la mesure du champ de pression se développant sur l'aube confirme cette hypothèse [5]. Elle illustre, en outre, le rôle prépondérant joué par le bord d'attaque où l'amplitude des fluctuations de pression augmente avec le taux de turbulence.

c) Modélisation du bruit émis
Les profils d'aubes ont, en général, une épaisseur suffisamment faible, comparée à leur corde (e/c < 12 %), pour qu'elles puissent être assimilées à une plaque plane. Dans ce cas, on peut utiliser les théories de l'aérohydrodynamique instationnaire pour prédire, à partir du spectre de vitesse de l'écoulement ambiant, la distribution de pression induite sur le profil par la turbulence. L'utilisation de l'équation de Fowcs-Williams et Hawking permet ensuite d'en déduire le bruit rayonné. Cette démarche a été utilisée par Goldstein [2], Amiet [6] et Arbey [5] avec succès. La distribution de pression induite sur l'aube ainsi que le spectre du bruit rayonné sont correctement prédits, pour des fréquences inférieures à 1.5 kHz, comme le montre la figure 1.

3 - Cas d'un obstacle non profilé

a) Écoulement sain
Le spectre du bruit rayonné par un cylindre circulaire placé dans un écoulement sain (donc le nombre de Reynolds basé sur le diamètre D est compris entre Re = 60 et Re = 2 000), correspond aussi en grande partie à la fréquence harmonique 2f_s. Ceci est illustré sur la figure 2 pour un obstacle de diamètre D = 16 mm et une vitesse d'écoulement de U = 29 m/s (Re = 3.2 x 10^5).

Le mécanisme associé à cette émission sonore est bien connu [2], [7]. À l'arrièrè de l'obstacle se développe une allée tourbillonnaire qui induit sur l'obstacle un champ de pression non homogène. Comme le montrent les mesures de Bruun [9] effectuées sur le périphérie d'une section droite du cylindre, la spectre de pression en sagement sur la partie arrière de l'obstacle est dominé par un pic à la fréquence f_s. L'amplitude du pic croît avec l'angle d'observation a et est maximale pour a = 90°. Le comportement analogue est observé pour la composante large bande du spectre d'E. Les niveaux de fluctuation croissent avec l'angle a. Enfin, le spectre en mesuré à l'aval du cylindre (a = 180°) est dominé par un pic à la fréquence 2f_s. Ces fluctuations de pression génèrent sur l'obstacle une portance fluctuante de fréquence f_s et une trainée fluctuante de moindre amplitude, de fréquence 2f_s. Ces grandeurs fluctuantes sont à leur tour responsables des pics observés dans le spectre du bruit rayonné pour ces 2 fréquences.

Précisons qu'un mécanisme analogue à long temps a été utilisé pour expliquer l'émission sonore de corps profilé avant que [4] ne propose le modèle rappelé dans la première partie. Cette confusion peut s'expliquer de la façon suivante. En écoulement sain, bien que les mécanismes de production du champ de pression soient différents (instabilités de couches limites pour l'obstacle profilé, échappement tourbillonnaire pour l'obstacle cylindrique), dans chacun des cas les champs de pression induits sont localisés dans la région aval des obstacles.

b) Écoulement incident turbulent
Plaçons nous dans une plage de nombre de Reynolds suffisamment faible pour ne pas être affecté par la transition des couches de cisaillement. Alors, une augmentation du taux de turbulence amène à traduire, sur le spectre de pression mesuré sur la face avant...
de l'obstacle (|α| < 90°), essentiellement par un enrichissement des composantes basses fréquences. Cet effet devient prépondérant près du point d'arrêt (|α| < 30°) où le spectre de pression est relié au spectre de la pré turbulence conformément à la théorie de la déformation rapide de la turbulence [11] et comme le confirment les expériences de Huot [12] effectuées à l'aide d'un cylindre à base carrée. Au point d'arrêt (α = 0°) l'amplitude du pic de Strouhal décroît lorsque le taux de turbulence augmente. Il disparaît complètement pour des taux supérieurs à 10%. Les points situés à l'arrière de l'obstacle (|α| > 90°) sont peu affectés par la pré turbulence.

Notons à nouveau, la similitude entre le cas d'un obstacle profilé et celui d'un obstacle non profilé. Les fluctuations de pression induites par la pré turbulence sont dominantes dans la région amont de l'obstacle.

c) modélisation du bruit émis

A partir des résultats de Bruun, on peut considérer, en première approximation, que la contribution du sillage au spectre de pression (pic de Strouhal \( f_s \)) est statistiquement indépendante de la contribution de la pré turbulence (spectre large bande).

Pour prédire le niveau sonore émis à la fréquence de Strouhal \( f_s \), différentes approches sont possibles [8]. Celle de Goldstein [2] consiste à assimiler la portance fluctuante à un dipôle puis à utiliser l'équation de Frowen-Williams et Hawkings. L'accord avec les résultats expérimentaux est satisfaisant [2].

Pour modéliser la contribution large bande du spectre de bruit rayonné, on peut avoir recours à une démarche analogue à celle utilisée par [5] et [6] dans le cas d'un obstacle profilé. La théorie de la déformation rapide de la turbulence permet d'obtenir, à partir du spectre de vitesse de l'écoulement incident, le spectre de pression en développant sur la partie amont de l'obstacle (|α| < \( \pi/2 \)). En supposant que seuls ces points participent au bruit rayonné (hypothèse confirmée par les travaux de Bruun), l'utilisation de l'équation de Frowen-Williams et Hawkings fournit alors le spectre du bruit émis.

**Bibliographie**


Figure 1 : Spectre du bruit rayonné par un profil d'aube placé dans un écoulement sain (\( u'/U < 3 \% \)) puis perturbé (\( u'/U = 12 \% \)).

Figure 2 : Spectre du bruit rayonné par un cylindre circulaire placé en écoulement sain.
L'ETUDE DU BRUIT DES LIGNES ELECTRIQUES A E.D.F.

J. Delcambre, A. Petitjean

Département Acoustique, E.D.F. - DER, 1 avenue du Général de Gaulle 92141 Clamart, France

INTRODUCTION

Sur les 38,000 km de lignes très haute tension (225 kV et 400 kV) en courant alternatif que compte la France, une dizaine de situations jugées bruyantes par le voisinage ont été soumises au Département Acoustique d'Electricité de France au cours des deux dernières années. Pour les traiter, un large programme de recherche, dont nous présentons les grandes lignes, est actuellement engagé.

Nous distinguerons le bruit électrique d'effet couronne, les bruits d'oliens et les résonances de cavité.

1. METHODES D'INVESTIGATION

Au-delà de mesures ponctuelles, un effort particulier de compréhension, tant expérimental que numérique, est mené au sujet de la génération même de ces bruits. Dans le cas où l'étude nécessiterait la prise en compte de la propagation, E.D.F. dispose des programmes provisionnels intégrant les effets de sol et les conditions météorologiques.

L'analyse sur site est facilitée avec une chaîne de surveillance longue durée qui enregistre les données acoustiques et météorologiques et permet des analyses statistiques (distributions statistiques, régressions multiples, etc.).

Les analyses spectrales des bruits de lignes permettent, en général, de distinguer le bruit électrique à haute fréquence des bruits d'oliens à basse fréquence.

2. BRUIT D'EFFET COURONNE

2.1 Mécanismes physiques

Le bruit électrique d'effet couronne a pour origine l'ionisation de l'air au voisinage du conducteur. L'équation générale de génération et propagation permet d'analyser l'influence acoustique des différents mécanismes physiques de l'effet couronne (transfert de masse, de quantité de mouvement et d'énergie).

2.2 Calcul prévisionnel

A partir des travaux expérimentaux [1], le bruit d'un faisceau peut être évalué par une relation du type :

\[ N'_{db(A)} = N_0 + 10 \log 4 \log r + 10 \log n - 10 \log D \]

où \( g \) est le gradient de tension superficielle, \( r \) la rayon des conducteurs, \( n \) le nombre de conducteurs et \( D \) la distance ligne observateur.

Cependant, l'état de surface du conducteur et les conditions météorologiques peuvent entraîner des variations de 20 dB, d'où des niveaux dépassant 50 dB à une vingtaine de mètres d'une ligne 225 kV à 2 terres. Les limites de la méthode précédente justifient donc le développement de nouveaux moyens d'analyse et de calcul.

3. BRUIT EOLIEN DES CABLES

3.1 Origin physique

Le vent soufflant sur les câbles provoque des bruits pouvant atteindre 70 dB(A) à 20 mètres de la ligne. Le détachement alterné de tourbillons engendre des efforts fluctuants à des fréquences harmoniques de la fréquence de Strouhal f telle que le nombre de Strouhal St = fо/ν reste de l'ordre de 0,2.

On examinera l'influence de la turbulence du vent et celle du mouvement des câbles (mesures laser).

3.2 Essais en soufflerie

Un tronçon de câble ASTER 570 de longueur 60 cm a été soumis à des vitesses comprises entre 10 et 25 m/s dans la soufflerie anéchoïque de l'Ecole Centrale de Lyon.

La fréquence du pic de Strouhal croît proportionnellement à la vitesse avec un décrochage à 19 m/s dû à la rugosité du câble.

Des cordons enroulés en spirale autour du câble (diamètre 8 mm, pas 10 cm) permettent de réduire le niveau du pic de 10 dB et le niveau global de 7 dB(A) (cf [4] et [5]).

Fig 1: Efficiency of helical trip wires

- (1) Clean ASTER 570 cable
- (2) Cable with trip wire (diameter = 8mm, pitch = 10cm)
- (3) Cable with trip wire (diameter = 8mm, pitch = 5cm)

À la suite d'une plainte, l'étude de l'interaction des sillages de 2 câbles a été entreprise.

Des mesures en soufflerie montrent qu'un écartement des câbles de 600 mm permet de gagner 3 dB(A) par rapport à l'écartement de 400 mm. Ceci est complété par une modélisation numérique.

3.3 Modélisation numérique

L'écoulement est supposé uniforme et le câble immobile dans un premier temps. Selon la théorie de Curle, le bruit provient de la fluctuation de portance sur le câble. Celle-ci est calculée en résolvant les équations de Navier-Stokes autour d'un cylindre.

Le modèle en différences finies et incompressible Estet permet de visualiser les tourbillons de Karman et d'obtenir un nombre de Strouhal (St = 0,2) et un coefficient de portance (Cp = 0,6) corrects.
4.3 Autres matériaux
De même que les entretoises, les anneaux pare-effluves présentent des cavités qui peuvent potentiellement générer des sons purs ; l'obturation des trous existants lors de la mise en forme des anneaux constitue une solution radicale.
En ce qui concerne les supports, ils semblent rarement à l'origine de bruits éoliens pour les types "chat" ou "trianon". Pour quelques autres cas, le bruit a été suspecté mais n'a pas pu être mis en évidence clairement.

5. CONCLUSION
Les bruits éoliens peuvent être évités ou réduits assez facilement. Par contre, le bruit d'effet couronne est plus difficile à réduire. Comptant tenu à la fois du développement prévu sur le réseau haute tension et du nombre de plaintes annuel, les outils présentés précédemment devraient permettre de prévoir et de maîtriser les bruits des lignes haute tension pour concevoir un réseau aussi silencieux que possible.

REFERENCES
[1] Gary C., Moreau M.
L'effet de couronne en tension alternative.
Eyrolles, 1976
Emission sonore d'un système de cavités périodiques ouvertes
1ère ICA, 1983
[3] Petitjean A., Esposito P.
Un premier calcul du bruit éolien d'une ligne électrique
Note EDF, HE/22 - 4807
Wind tunnel investigation of aeolian tones from rough surface cylinders
IUTAM Lyon, 7/1985
[5] Petitjean A.
Étude expérimentale du bruit éolien des câbles électriques
Note EDF, HE/22 - 1883
[6] Petitjean A.
Essais d'entretoises en soufflerie
Note EDF, HE/22 - 1881

4. BRUITS ÉOLIENS DE MATÉRIAUX DIVERS
4.1 Isolateurs
Les études antérieures ([2]) ont montré que le bruit des chaînes d'isolateurs est indépendant du matériau et provient d'un bouclage aéroacoustique sur le bord de la jupe et la première nervure. Celui-ci peut être amplifié par une résonance acoustique des cavités interisolateurs couplées, d'où des niveaux très importants (jusqu'à 70 dB à 100 m). D'un point de vue pratique, on supprime ce bruit en intercalant un isolateur d'un type différent (par exemple isolateur à jupe longue) tous les 4 ou 5 isolateurs.

4.2 Entretoises
Certaines entretoises engendrent également des bruits éoliens dont le mécanisme est du type résonance de cavités ou bouclage aéroacoustique. C'est pourquoi les spécifications techniques des entretoises comprennent un essai aéroacoustique en soufflerie anéchoïque destiné à éviter ces sons purs.
LOUDNESS AND ANNOYANCE OF THE ACOUSTIC NOISE PRODUCED BY CORONA DISCHARGES IN A.C. POWER LINES.

M. Fiorina (1), R. Piazza (2), F. Rosa (3)

(1) ENEL - Centro Ricerca di Automatica
Via A. Volta, 1 - 20093 Cologno Monzese, Milano, Italy

(2) Istituto di Fisiologia dell'Università di Torino - Viale Mattioli, 25 - 10125 Torino, Italy

(3) ENEL - Centro di Ricerca Elettrica
Via A. Volta, 1 - 20093 Cologno Monzese, Milano, Italy

INTRODUCTION

In the framework of the 1000 kV Project, systematic research on the acoustic noise produced by corona on bundled conductors for UHV lines has been conducted [1].

Moreover, experimental investigations on the physiological and psychological aspects of the noise produced by a.c. corona discharges have been done [2] and are still in progress, as it is characterized by physical parameters that distinguish it from other types of noise. The typical complex spectrum consisting of a few pure tones and continuous components extending up to ultrasonic frequencies, the stationarity, the large variability with atmospheric conditions raise the question whether it would be legitimate to apply to the corona noise the same frequency weighting methods and criteria for a correct assessment of acceptable levels considered suitable for other more common noises: a direct knowledge of the loudness and annoyance characteristics of the corona noise is actually required. To give a general view of the studies conducted in this field, the paper reports a summary of the research on the loudness already performed and the results of a first series of annoyance tests.

LOUDNESS

With the co-operation of the Istituto Elettrotecnico Nazionale Galileo Ferraris (IENNF) and the University of Turin and making use of a certain number of samples of corona noise recorded during tests on conductor sections set up in the IENNF's anechoic chamber, experimental investigations have been conducted aiming at the assessment of the loudness of the noise, by listening tests performed in anechoic chamber with 20 subjects of normal hearing [2].

Three main aspects of the problem were investigated.
1) The influence of the shape of the frequency spectrum, especially of the level of the 100 Hz component, on the loudness of corona noise.
2) The loudness scale of corona noise, i.e., the variation of the loudness as a function of the sound pressure level of the noise.
3) The comparison of the loudness of corona noise with that of other types of noise.

As regards the first aspect, the results of the listening tests indicate that the shape of the frequency spectrum (slope of the continuous component, relative level of the 100 Hz tone) has an important effect on the loudness of the noise. The usual weighting methods (A, B, etc.) are generally not strictly related to the loudness of the noise throughout the range of corona noise frequency spectra. In particular, for the same dB(A) level, the loudness increases when the 100 Hz component increases, as shown in Fig. 1, that summarizes the main results obtained from 11 corona noises of different spectral composition, of which one was assumed as a reference (R). Fig. 1 also indicates possible correction factors to be applied to the dB(A) level as a function of the difference between the 100 Hz tone and the dB(A) levels and shows how in extreme cases the same loudness can be assigned to noises - (6) and (7) - whose levels differ by 8-9 dB(A).

![Fig. 1 - Effect of the 100 Hz pure tone level on the loudness of corona noise. $\Delta L_A$ are the differences between the level of a reference noise (R) and the noise (i) for which 50% of the listeners considered the reference noise louder than the other.](https://www.example.com/image)

As regards the second aspect, the scale of loudness of the corona noise (i.e., the variation law of the loudness as a function of the noise level) was obtained by applying the loudness ratio estimation method, with ratios of 2 and 1/2 (corresponding to doubling and halving of the loudness). The results of the listening tests indicate that a doubling or halving of the loudness is obtained, on the average, for an increase or decrease in the sound pressure level of 6.5 or 7.4 dB, respectively.

In Fig. 2 the average value of 7 dB is compared with the corresponding value of 10 dB, characteristic of the 1000 Hz pure tone (slopes of the two straight lines). When the absolute values are considered, the same Fig. 2 also shows the results of the loudness comparison between corona noise and the 1000 Hz pure tone.

ANNOYANCE

In the listening tests for the estimation of the annoyance, the noise with the spectral composition given in Fig. 3 was employed. This noise, obtained in light rain conditions nearby the ENEL'S Susveto test line, shows the mean characteristics of noise of possible 1050 kV three-phase lines. This noise was recorded on magnetic tape with a Nagra IV SJ tape recorder with input from a precision sound-level meter (B&K 2202, microphone B&K 4165). From this recording,
Fig. 2 - Loudness of corona noise compared with that of the 1000 Hz pure tone.

A new magnetic tape was prepared containing 16 sections, each consisting of a 20 s stimulus interval followed by a 15 s silent one, in such a way as to repeat the same portion of the original tape at different levels, randomly covering in steps of 2 dB or multiples the range -16 to +16 dB from a reference level.

![Frequency spectrum of corona noise](image)

Fig. 3 - Frequency spectrum of the corona noise utilized in the annoyance tests.

The listening tests were performed in the anechoic chamber, where the stimuli were presented to groups of 6-7 subjects (young university students) up to a total number of 16-20.

For each section the listeners expressed their judgement on the basis of an empirical annoyance scale, defined as indicated in Fig. 4.

The results of a first series of tests, expressed in terms of mean values of annoyance rating, are given in the same figure.

The range of variation of the individual ratings around the mean value was generally quite large (standard deviation: 0.8-1.2 steps). In addition, the figure shows that the annoyance rating is affected by the reference level. All these facts indicate that a definite assessment of the annoyance for a given sound pressure level of corona noise appears to be very difficult, as already found by other experimenters for other types of noise.

![Mean annoyance rating of corona noise for three different reference levels](image)

Fig. 4 - Mean annoyance rating of corona noise for three different reference levels: 41, 48.5, 55 dB(A).

In the case of Fig. 4, the differences in rating for different reference levels are probably imputable to the procedure adopted, that seems to have not sufficiently opposed the subjects' propensity for giving relative rather than absolute judgements; in fact the points referring to the three different levels are arranged along almost parallel lines, at least in the median part of the diagram.

Other methods are being experimented to try to minimize the main causes of variability.

In addition, the future research program will consider the study of the influence of the corona noise spectrum characteristics on the annoyance rating and the definition of general criteria for assessing acceptable noise levels, taking into account the statistical distribution of the noise.

REFERENCES


AUDIBLE NOISE FROM AN ONTARIO HYDRO 500 KV TRANSMISSION LINE

J.J. Kowalcwski, J. Reichman
Ontario Hydro Research Division
800 Kipling Avenue, Toronto, Canada, N6Z 5S4

Audible noise from overhead transmission lines has not been a problem in Ontario Hydro, considering there are about 1700 km of 500 kV line in service throughout the province. This may be attributed to a combination of noise control at the source by design and routing of the lines so as to minimize noise impact on potential recipients. In order to verify the design criteria and better understand the noise impact from these particular lines, some field measurements have been collected.

Design Considerations

The present design criteria used by Ontario Hydro for limiting audible noise from new transmission lines is 55 dBA – one hourly $L_{eq}$. This applies at any point on the premises of a person within 30 m of a dwelling or of a campsite/1/. In extreme cases the limit could apply at a point just outside the right-of-way boundary.

The predicted noise from a given line design is derived from the following empirical relation:

$$AN = 100 \log g + 40 \log d - 10 \log D - 77.2$$

where: "AN" is the overall level in dBA, expected during substantially wet conditions. It applies over short distances, up to about 60 m, where excess attenuation due to air absorption can be neglected.

"g" is voltage gradient, kV/cm

"d" is subconductor diameter, cm

and "D" is radial distance to outside phase, m

This relation is in good agreement with other methods reported in the literature/2/.

According to this formula the maximum noise levels expected at 15 m lateral distance from the outer phase of a single and double circuit lines (at 550 kV) are 53 and 55 dBA respectively.

Field Verification

Although the predicted levels were confirmed in the past by "spot" measurements, a more comprehensive test program was undertaken recently to verify the results. The general arrangement of the test line and microphone positions are shown in Figure 1.

Two outdoor microphones, at 15 m and 60 m laterally from the closest phase, were used to simultaneously record the analog sounds on a two channel tape recorder. This information along with meteorological data were collected in-situ automatically during corona noise occurrences using a field monitoring facility shown in Figure 2.

Occasionally, measurements were also made manually using back-up equipment to check the accuracy of the automatic system. Data collected by both systems were analyzed in the laboratory using the set up shown in Figure 3.

The two microphone method was considered essential for differentiating corona noise and extraneous environmental noises.
Results

The analog recordings facilitated identification of noise samples which were predominantly due to corona from the test line. The energy equivalent level, over the duration of each sample, was then calculated and tabulated by computer. An example is shown in the Table.

<table>
<thead>
<tr>
<th>DATE</th>
<th>TIME</th>
<th>DATA SOURCE (m)</th>
<th>WEATHER CONDITIONS</th>
<th>LEQ</th>
<th>OVERALL &quot;L EQ&quot;</th>
<th>PRECIPITATION</th>
<th>WIND SPEED</th>
<th>WIND DIRECTION</th>
<th>TEMPERATURE</th>
<th>RELATIVE HUMIDITY</th>
<th>NUMBER OF AVERAGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>04-11-83</td>
<td>14:35</td>
<td>AUT</td>
<td>9</td>
<td>15</td>
<td>150</td>
<td>80</td>
<td>0.2</td>
<td>68</td>
<td>75</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>04-11-83</td>
<td>14:15</td>
<td>AUT</td>
<td>9</td>
<td>15</td>
<td>150</td>
<td>80</td>
<td>0.2</td>
<td>68</td>
<td>75</td>
<td>80</td>
<td>10</td>
</tr>
<tr>
<td>09-11-83</td>
<td>23:11</td>
<td>AUT</td>
<td>5</td>
<td>30</td>
<td>300</td>
<td>100</td>
<td>5</td>
<td>45</td>
<td>40</td>
<td>85</td>
<td>10</td>
</tr>
<tr>
<td>09-11-83</td>
<td>22:17</td>
<td>AUT</td>
<td>5</td>
<td>30</td>
<td>300</td>
<td>100</td>
<td>5</td>
<td>45</td>
<td>40</td>
<td>85</td>
<td>10</td>
</tr>
<tr>
<td>10-11-83</td>
<td>22:14</td>
<td>AUT</td>
<td>4</td>
<td>10</td>
<td>200</td>
<td>100</td>
<td>5</td>
<td>45</td>
<td>40</td>
<td>85</td>
<td>10</td>
</tr>
<tr>
<td>10-11-83</td>
<td>22:13</td>
<td>AUT</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>5</td>
<td>45</td>
<td>40</td>
<td>85</td>
<td>10</td>
</tr>
<tr>
<td>10-11-83</td>
<td>22:09</td>
<td>AUT</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>5</td>
<td>45</td>
<td>40</td>
<td>85</td>
<td>10</td>
</tr>
</tbody>
</table>

* Energy equivalent sound pressure level.

This example shows typical levels measured during periods of light to heavy rain. At the 15 m location, corona noise levels, in short term levels, were within the predicted level of 51 dBA (calculated for a line voltage of 525 kV existing at the time). At the remote location the sound level was up to 4 dB lower.

Time histories of the two microphone signals, such as in Figure 4, gave an indication of corona noise recordings that may have been contaminated by background noise from other sources – ie rain, wind, insects, etc. It was observed from this data that corona noise levels during rain vary closely with rain intensity.

Acknowledgement

The authors wish to thank the Ontario Hydro Design and Development Division - Transmission for their support of this study.

References


Conclusion

Based on this study it was found that audible noise produced by a typical Ontario Hydro 300 kV transmission line is within the predicted level and, therefore, the design criteria. The maximum levels at the right-of-way boundary are below 55 dBA, one hourly Leq, and often are lower than other environmental noises such as rainfall itself.
EFFECT OF AUDIBLE NOISE ON UHV POWER LINE DESIGN

R. Cortina (1), F. Rosa (2)

(1) ENEL - Centro Ricerca di Automatica
Via Volta, 1 - 20093 Cologno M., Milano, Italy
(2) ENEL - Centro Ricerca Elettrica
Via A. Volta, 1 - 20093 Cologno M., Milano, Italy

INTRODUCTION

The knowledge of the phenomena associated with the generation and propagation of acoustic noise produced by A.C. power lines has reached substantial improvement in the last years. In various countries, involved in UHV transmission, general criteria to assess the influence of the line parameters on the noise level and to define the most suitable methods to express the human response to corona noise have been developed. In Italy, in particular, in the framework of the 1000 kV Project, particular attention has been devoted to the problems of corona acoustic noise, by testing various conductor configurations on the 1 km test line and on the cage of Suvereto test station [1] and by conducting loudness and annoyance listening tests on corona noise samples [2].

The paper intends to show, on the basis of the present knowledge on the problem, how the audible noise level control affects the UHV power line design, taking into account technical and economical impacts.

PREDETERMINATION OF CORONA ACOUSTIC NOISE

Corona generation, and in particular acoustic noise, of a conductor bundle is mainly affected by the voltage gradient distribution on and around the conductors of the bundle and by surface conditions [2].

For the most common case of symmetrical bundles (subconductors equally spaced along a circumference) the overall bundle noise generation may be simply expressed as a function of the average value of the maximum voltage gradients of the subconductors, \( g_{\text{max}} \), the subconductor diameter \( d \), the number of subconductors in the bundle \( n \).

As regards surface conditions, for practical purposes, the following conditions may be considered:
- new and aged conductors;
- dry or wet (fog, dew, after rain) conductors, rain, snow, ice, etc.

The noise is higher for new conductors than for aged conductors and reaches the maximum value in heavy rain conditions. However, the new conductor condition is limited to a few months of line operation and the noise in heavy rain is masked by the noise of the rain itself.

For UHV conductor bundles, the majority of the researches are conducted in cages using artificial heavy rain and measuring the acoustic noise in this condition and during the drying period of the conductors (wet conductor conditions). The acoustic noise is generally expressed in dB(A), but recordings of frequency spectra in various conditions are also made to better characterize the noise.

The following procedure should be adopted to obtain the acoustic noise level to be considered for line design.

1) Determination, starting from the above measurements, of the acoustic power of each phase conductor in heavy rain conditions, for new and aged conductors;
2) Application of reduction factors, as a function of voltage gradient, subconductor diameter and number of subconductors, to be applied to the heavy rain values, in order to obtain the corresponding levels for other conditions (rain of other intensities, wet conductor, etc.);
3) Determination of the acoustic pressure at a given distance from the line on the basis of the acoustic power of each phase conductor, for the various surface conditions;
4) Determination of the annual statistical distribution of the acoustic noise, taking into account the frequency of occurrence of rain intensities and of the other conditions (dry, wet conductor, etc.);
5) Determination of "equivalent" noise levels, for indoor and outdoor conditions, taking into account the above statistical distribution, background noise (in particular, rain), frequency spectrum, noise attenuations of building walls, loudness and annoyance characteristics;
6) Comparison of the above "equivalent" levels with some reference design values.

With reference to item 1), various formulas have been set up in the last years. Recently, on the basis of the research conducted within the 1000 kV Project, the following formula was established for symmetrical bundles, in heavy rain condition:

\[
A = -14 - 576 g_{\text{max}} + 37 \log d + 10 \log n
\]

where \( A \) = generated acoustic power of a bundle (dB(A) above 1 \( \mu \)W/m)
\( g_{\text{max}} \) = bundle average maximum voltage gradient (kV/m/m)
\( d \) = subconductor diameter (mm)
\( n \) = number of subconductors.

The above formula refers to aged conductors; for new conductors, the influence of the line parameters, and especially of the gradient, is generally less pronounced.

As regards the reduction factors to obtain the noise levels in different conditions than in heavy rain, a complete knowledge of the problem is not yet available. From the qualitative point of view, the influence of rain intensity is more pronounced for lower voltage gradients, lower subconductor diameters, lower numbers of subconductors, aged conductors and asymmetrical bundles. Fig. 1 shows, as an example, a diagram referring to an aged 8 conductor bundle in symmetrical and asymmetrical configuration: in the same figure the heavy rain and wet conductor levels are also reported.

As regards the criteria for the evaluation of the overall line noise starting from the noise of each bundle and the determination of the annual statistical distributions, there are no particular problems. In practice, however, the statistical analysis may be considerably simplified considering only the noise
Fig. 1 - Influence of rain intensity on audible noise, obtained from continuous recording under the Suvereto test line on R×21.5 mm asymmetrical and symmetrical bundles.

The level corresponding to an average rain intensity.

Finally, the criterion for the evaluation of the "equivalent" noise level requires further investigation, especially with reference to the definition of loudness and annoyance of corona noise[2].

LINE PARAMETERS AFFECTING ACOUSTIC NOISE

As already mentioned, the acoustic noise of a symmetrical bundle is affected by voltage gradients, subconductor diameter and number of subconductors; voltage gradients depend, on their turn, for a given voltage, by the bundle geometry, the distance between adjacent phases and the height above ground. The effect of the various parameters is different: as an example, Fig. 2 shows that the variation of the number of subconductors in the bundle is more effective than that of the subconductor diameter and of the distance between phases.

Fig. 2 - Effect of some parameters on the wet conductor audible noise of 150 kV lines at 15 m from the lateral phase.

Fig. 3 - Overall cost of optimal solutions of a 150 kV line, as a function of the design wet conductor audible noise level at the edge of the right-of-way, for two values of electric field at ground and of right-of-way cost.

The final choice of the line parameters has to take into account the technical requirements (insulation clearances, electric field at ground, etc.) and economical impact. A systematic study on this problem is reported in [2]. Fig. 3, taken from this report, gives an indication of the cost variation of an UHV line as a function of the design acoustic noise level.

Particular methods to reduce acoustic noise (asymmetrical bundles – see Fig. 1; – tubular conductors; conductor surface treatment) have been developed. Their application, however, has to be limited to solve local problems, since it generally involves some drawbacks (worse mechanical behaviour, higher losses, etc.).

REFERENCES


We suggest a simple model for assessment of annoyance from road traffic noise. The model is based on Leq and Lmax and the percentage of heavy vehicles. The validity of the model has been demonstrated by laboratory experiments.

Introduction

In a previous paper we have shown that a threshold model based on contributions to the annoyance only from those periods when the noise level exceeds a specific threshold is a better and more logical way to assess noise annoyance than using the regular equivalent level. [1].

This new noise index "equivalent level above a threshold" Leq, is computed basically in the same way as the regular Leq level. Lmax, with the exception that those periods when the noise level drops below a predetermined threshold are omitted.

A relationship between annoyance and LTeq was presented by us in [2]. We used a subjective estimate method with a 10-point annoyance scale ranging from "not annoyed at all" to "extremely annoyed". Figure 1 shows annoyance versus LTeq based on a laboratory experiment. The LTeq index gives a better fit to our data points and the percentage Lmax in his comments on the method. Schulte [3] writes: "the coefficient of the standard error for the curve is much more plausible looking fit to the data points than the Lmax solution."

The Threshold Concept

We need to establish a threshold below which a particular noise event is not considered annoying. This threshold may not necessarily coincide with the detection threshold, but more likely is somewhat higher. Obviously this threshold will depend on the number of variables such as activity, expectations, the presence of "noisier" sound etc.

We have no firm bases for choosing these thresholds but temporarily we will suggest the following:

Rest and sleep: 25 dBA
Quiet day time: 40 dBA
Busy day time: 55 dBA

Further investigations may prove that different thresholds will yield even better results. This, however, will not alter the fundamental procedure of our method.

If we choose a low threshold most of the noise will contribute to Leq as the noise seldom drops below the threshold. LTeq will thus have the same value as Leq.

At increasing threshold levels an increasing portion of the noise will be omitted for the Leq calculations. LTeq will thus gradually decrease and drop infinitely for threshold levels above the maximum noise level, corresponding to a situation when the noise is completely "masked" and consequently not annoying at all. Figure 2 shows the difference between Leq and LTeq for road traffic noise as a function of the threshold level expressed in dB below Lmax. This relationship is different for different percentage of heavy vehicles.

Annoyance assessment

For a given traffic situation we measure LTeq and Lmax and calculate the percentage of heavy vehicles. We choose the appropriate threshold level and enter these values Figure 2 to find the difference between Leq and LTeq. Figure 1 is used to find the corresponding annoyance score.

For noise planning purposes we may use a noise prediction model to find LTeq and estimate a difference of 15 dBA between Leq and Lmax.

Experiment validation

We have carried out a small laboratory experiment to validate the model. We presented recorded traffic noise to six test subjects. We used two different traffic situations with 201 and 51 heavy vehicles respectively. The traffic noise was presented at 5 different listening levels, 38, 44 and 50 dBA. We used a constant background noise to create an artificial threshold as to noise sources were present in the laboratory. This procedure is explained in detail in [2].

The procedure for calculating the correction factors for LTeq is shown in Figure 3 and our annoyance scores are presented in Table 1 and Figure 4.

### Table 1

<table>
<thead>
<tr>
<th>Leq</th>
<th>Lmax</th>
<th>A</th>
<th>Leq</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>53</td>
<td>33.5</td>
<td>3.0</td>
<td>30.0</td>
</tr>
<tr>
<td>44</td>
<td>59</td>
<td>47.5</td>
<td>4.1</td>
<td>40.0</td>
</tr>
<tr>
<td>50</td>
<td>65</td>
<td>49.5</td>
<td>5.6</td>
<td>49.0</td>
</tr>
</tbody>
</table>

Annoyance scores and noise levels.

Conclusions

Although the regular equivalent noise level, Leq, seems to correlate fairly will with subjective assessments of noise annoyance, this noise index often fails in describing irregular noise situations especially at low and moderate noise levels. Similarly Leq also seems to be less fit in describing the effect of small changes in a specific noise situation.

Our model, with a modified equivalent level, LTeq, possesses much the same properties as the Leq-model, but in addition the LTeq-model gives a plausible explanation of the relatively large differences in annoyance that sometimes may be found for different traffic noise situations with almost identical Leq-levels.

In this article we have outlined a simple method for traffic noise annoyance assessment. However further refinements need to be carried out before the method may be applied to any traffic noise situation in general.
References


Figure 1
Annoyance versus $L_{T_{eq}}$

Figure 2
The difference between $L_{T_{eq}}$ and $L_{T_{eq}}$ for different traffic (5% heavy vehicles) as a function of threshold level below $L_{max}$

Figure 3
Calculation of $L_{T_{eq}}$ for the lab experiment

Figure 4
Annoyance scores from the lab experiment
RESIDENTIAL NOISE EXPOSURE AND REACTION OF INHABITANTS AROUND ROADSIDE AREA

K. Kuno, K. Ohara, K. Takeda and Y. Mishina*

Electrical Eng., Nagoya University, Chikusaku, Nagoya 464 JAPAN
*Daldo Inst. of Tech, Daldo Kouyou DAlgaku, Minakiku, Nagoya 457 JAPAN

1. INTRODUCTION

Measurements of residential noise exposure over 24 hours are made in Nagoya City. Associated information about residence and reactions of inhabitants to noise environment are also obtained. The survey has been continuing since 1982. These data, including noise ratings calculated from raw data, have been stored in computer system (FACON M-382: Computation Center of Nagoya University) to construct a data bank system on daily noise environment. About 600 samples are stored now(1-3).

The content of data is listed in Table 1. In the present paper, based on these data we investigate noise environment relevant to trunk road. Attenuation Characteristics of residential noise exposure due to distance from trunk road are considered and the reactions of inhabitants to noise environment are also discussed with respect to trunk road. Then it is shown that the influence of trunk road in urban area is mainly restricted to 20 ~ 30 meters from it.

2. Residential noise exposure around trunk road

The measurement of residential noise exposure is made by using Leq meter (Matsushita VR-42048 or Rion NE-13A) which records Leq(10min.) for every ten minutes over 24 hours. The meter is settled to a place which represents outdoor noise environment of the residence. From the data Leq24 and other typical noise ratings are calculated (see Table 1). Dependence of these noise ratings on the distance from trunk road is shown in Fig.1. We can extract by inspection several prominent features which are summarized as follows.

(1) Residential noise exposure decreases very rapidly in about 20 meters from a road. In the distance LLeq,M, (also LLeq,E), LLeq,D, LLeq,N and LLeq24 decrease about 11 dBA, 8 dBA, 13 dBA and 9 dBA respectively. Level difference is especially large in the nighttime.

(2) Behind the above roadside area exposure level decreases very slowly and shows somewhat flat characteristics. From the viewpoint of noise exposure the influential region of trunk road is about 20 meters or so and the influence is dominant in the nighttime.

3. Reactions of inhabitants

Reactions of inhabitants to noise environment are obtained by interview with questionnaire shown in Table 2. Scaling of the reactions is made by using Likert's method. The average score for each reaction is plotted against the distance from trunk road. Prominent features of the curves are summarized as follows.

<table>
<thead>
<tr>
<th>Table 1 Questionnaire to inhabitants for their noise environments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>[IA]</strong> How do you feel noise level around your residence?</td>
</tr>
<tr>
<td>Answer: (1) loud (2) medium (3) low</td>
</tr>
<tr>
<td><strong>[IC]</strong> Are you annoyed by it?</td>
</tr>
<tr>
<td>Answer: (1) very annoyed (2) annoyed (3) not so annoyed (4) not annoyed</td>
</tr>
<tr>
<td><strong>[IB]</strong> Do you feel noisy around your residence?</td>
</tr>
<tr>
<td>Answer: (1) very noisy (2) pretty noisy (3) noisy (4) not so noisy (5) quiet (6) very quiet</td>
</tr>
<tr>
<td><strong>[ID]</strong> How do you think about the noise?</td>
</tr>
<tr>
<td>Answer: (1) should be abated (2) desirable to be abated (3) pay no much attention (4) pay no attention</td>
</tr>
<tr>
<td><strong>[IIA]</strong> How do you feel indoor noise level?</td>
</tr>
<tr>
<td>Answer: (1) loud (2) medium (3) low</td>
</tr>
<tr>
<td><strong>[IIC]</strong> Are you annoyed by it?</td>
</tr>
<tr>
<td>Answer: (1) very annoyed (2) annoyed (3) not so annoyed (4) not annoyed</td>
</tr>
</tbody>
</table>

![Fig.1 Relationships between residential noise exposure and distance from trunk road.](image-url)
Table 2 List of acquired data

[ row data ]
- $L_{Aeq}(10\text{min.})$ over 24 hours
- mesh code (= location of residence)
- attributes of residence and inhabitance
- land use
- traffic condition around residence
- noise sources
- reaction of inhabitants to residential noise environment
- sleep disturbance
  etc.

[ secondary data: (noise ratings) ]
- $L_{Aeq24}$, $L_{Adn}$, $L_{Aeq}$
- $L_{AeqM}$ : $L_{Aeq}$ (6:00-8:00)
- $L_{AeqD}$ : $L_{Aeq}$ (8:00-19:00)
- $L_{AeqE}$ : $L_{Aeq}$ (19:00-22:00)
- $L_{AeqN}$ : $L_{Aeq}$ (22:00-6:00)

4. Acknowledgements
The authors are obliged to Mr. A. Hayashi, Dr. Y. Oishi and the others who join and promote this survey. This work was supported in part by the Kajima Foundation's Research Grant.

References
2) K. Kuno et al. "Study on noise environment of residence in urban area."
Inter Noise 84 (1984.12) pp.953-958
3) K. Takeda et al. "Analysis of residential noise exposure patterns and scaling of reaction of inhabitants to noise environment."

(1) Response scores in and around residence decrease monotonically with distance from road.
(2) Each response score decreases rather rapidly within the roadside area limited in about 20 meters or so. The score become flat behind this area. The characteristics are clear in the response corresponding to the feeling on noise level in and around residence.

4. Conclusion
Influence region of road traffic noise in urban area is considered based on residential noise exposure and reactions of inhabitants. The results show that the serious effect is restricted within the roadside region nearby 20 meters or so.

Fig.2 Response of inhabitants to residential noise exposure around trunk road.
Response score is based on Likert’s scale.
ROAD TRAFFIC NOISE AND SLEEP. LONG-TERM EFFECTS, CRITICAL LOAD.

Barbara Griefahn
Institute for Occupational Health, University of Dusseldorf, Moorstr. 5, 4000 Dusseldorf, Fed. Rep. Germany

INTRODUCTION

The problem of noise-induced sleep disturbances increased during the last decades all over the civilized world. The main sources are neighbourhood, industry, and traffic.

Industrial noises and noises from the neighbourhood vary extremely with regard to intensities and frequencies, to the temporal distribution of noise peaks, and to the content of information and thus limits are difficult to determine. According to this a large number of different techniques is necessary for attenuation and they are often difficult to establish.

Road traffic noises are most often accused to cause sleep disturbances. Compared to other sounds this type of noise is relative uniform and that makes it easier to determine upper limits. In addition several countermeasures can be realized at low costs:
- streets can be opened in only one direction,
- maximum speed can be reduced,
- traffic lights can work according to need,
- the transit can be prohibited totally or selectively (motorcycles, trucks).

INVESTIGATION ON LONG-TERM EFFECTS. FIELD EXPERIMENTS

A large number of publications indicate that environmental stimuli such as noise contribute to the multifactorial genesis of several diseases. This hypothesis is supported by many investigations where no (complete) habituation was found.

If this is true also for nocturnal noise a critical load has to be determined. Habituation can be studied best in the field. The disadvantage, however, is the large number of intervening variables masking the reactions especially in small samples. Therefore a joint European program was carried out where 4 teams participated (France, Federal Republic of Germany, Great Britain, the Netherlands).

Method

Overall 70 subjects, men and women, 18-65 years of age took part in the experiments. They were in a sufficient state of health and with respect to psychologic constitution within a normal range. All subjects had lived in streets with high traffic density for at least 1 year.

The sleep of these subjects were observed during 12-23 nights each. Altogether almost 1,000 nights were recorded. The technical equipment for the recording of physical and physiological variables was installed in the subjects' own sleeping rooms.

Throughout the nights noise levels, EEG and EOG were recorded continuously. In the evening and in the morning as well the subjects completed a short questionnaire, where they assessed the actual situation as well as their sleep qualitatively and quantitatively. Thereafter a visual 4-choice-reaction-time-test was executed for at least 10 minutes.

Recordings were made during noisy and quiet nights as well. This was realized by the insertion of an experimental phase. Some subjects, who normally slept with closed windows opened the windows during the experimental nights. For other subjects the sound pressure then was reduced by using earplugs, by fitting double glazing, or by sleeping in the rear of the house. Thus, all subjects slept under both relative noisy and relative quiet conditions. The mean equivalent sound pressure levels in the sleeping rooms varied from 27-44 dBA during quiet and from 42-52 dBA during noisy nights.

RESULTS AND CONCLUSION

From the results of all subjects combined it can be stated that noisy nights are accompanied by the following:
- a decrease of REM-sleep
- an increase of the number and total duration of intermittent wakefulness
- an increase of reaction time as well as an increase of errors in the morning compared to the evening before
- a decrease of sleep quality as assessed by the subjects

Thus, for all 3 types of variables (EEG and EOG, subjective assessment, psychomotor performance) significant alterations due to noise exposure were found, but they are not necessarily correlated to each other. Nocturnal noise may elicit disturbances to be recorded during sleep and/or to after-effects which can be observed the next day or even later.

The small extent of the reactions suggests that habituation probably takes place. But as there is no correlation between residence time (over 1 year at least) and the extent of the reactions one most conclude that habituation remains incomplete. People living for many years in noisy streets still react to their usual acoustical environment.

DETERMINATION OF A CRITICAL LOAD. CONTROLLED EXPERIMENTS

Because the results of the European sleep study are confirmed by several other field experiments, the definition of a critical noise load is indispensable.

In comparison to the problem of habituation, the large number of intervening variables in the field is much more important when limits have to be determined. This is why a field study is hardly to realize, even for several collaborating teams. Limits can be determined best by controlled experiments in the lab.

Method

Thirty-six subjects, 18 males and 18 females, 21-30 years of age were randomly
selected from some hundred students. They
were healthy and psychologically in a normal
range. They all spent 12 consecutive nights in the lab.

During test series a high-density road traffic noise was played back with 4 intensities. The equivalent sound pressure levels varied from 37 to 63.5 dBA and were altered randomly every 2 or 3 nights. The maximum levels as indicated by the 1% value were 6 dBA higher.

As in the preceding field study sound pressure levels, EEG and EOG were recorded continuously throughout the nights. The same short questionnaire was completed every evening and every morning. Thereafter a visual reaction-time test was executed for at least 10 minutes. A 4-choice reaction-time test was applied alternatively with a simple reaction-time test.

During the preceding week as well as during the week following the experiments the subjects filled out the same questionnaire just before going to bed in the evening and after getting up in the morning. In addition the sound pressure levels were measured during one night in the usual sleeping rooms at home.

Results and Conclusion

Both the reaction-time tests were not affected by the noise levels applied. Among the physiological sleep parameters only stage REM (dream sleep) was altered. It was abruptly reduced as soon as the equivalent sound pressure level exceeded 44 dBA in the sleeping room. Thereafter no further REM deprivation was found (fig. 1).

Contrary to this subjective assessment was significantly related to the sound pressure levels (fig. 2). With increasing intensities the subjects felt to sleep worse, sleep onset was more and more delayed, the total time of intermittent wakefulness increased as well as the number of awakenings recalled in the morning and the degree of tiredness.

However, it is not possible to draw a critical load from this systematic relationship only.

To determine limits it is first of all necessary to define relevant criteria and appropriate baseline data. The noise levels measured at home were correlated with the assessment of sleep at home. None of these correlations were significant. This was verified for several subgroups (subjects who reported relative poor sleep). The conclusion then is that noise at home does not disturb sleep.

Comparing subjective assessment at home and in the lab no difference was found between the usual situation and the lowest intensity in the lab, whereas noisier conditions resulted in significant poorer sleep. Thus the experimental condition with the lowest sound pressure is regarded as a baseline.

The regressions then calculated between the sound pressure levels and the assessment of sleep quality, the time for sleep onset, the number of awakenings, and the total time of intermittent wakefulness were significant on the 1% level.

Critical noise loads for the different criteria can be calculated by insertion of the baseline data into the appropriate function. They vary between 39.7 dBA and 42.6 dBA. From the viewpoint of preventive medicine it is reasonable to choose the minimum that is to say an equivalent sound pressure level of 40 dBA indoors.

Hints to the validity of this critical load are found in different investigations. But it is probably applicable only for high-density road traffic noise where the differences between the average sound level and the maximum levels remain beneath 10 dBA. The validity has to be proven or new limits have to be established for larger differences that means for lower traffic densities.

References
Grifahm, B.; 1985: Schlafverhalten und
Görgeschs. Enke, Stuttgart.
Juririen, A.A. et al.; 1983: An Essay in European
Research Collaboration. In: Rossi, G.
(ed): Noise as a Public Health Problem.
Milano, Italy, pp 965-971
MUTUAL RELATION AMONG ENVIRONMENTAL NOISE EVALUATION INDICES ON THE BASIS OF \( L_\text{eq} \) AND \( L_\text{x} \) EVALUATION INDICES AND ITS APPLICATION TO THE ACTUAL ROAD TRAFFIC NOISE

Yasu Kimura* and Mitsuo Ohta**

*Faculty of Engineering, Nihon University, Sanjo, Higashimura-cho, Sakuragawa City, 750-85, Japan
**Faculty of Engineering, Hiroshima University, Chitani, Saijo-cho, Higashihiroshima City, 724, Japan

INTRODUCTION

As is well-known, the noise evaluation indices, \( L_\text{x} \) and \( L_\text{eq} \), are very important in the field of noise evaluation and its applications. From the practical point of view, a general theory for the mutual relationship among several noise evaluation indices is very important to grasp the systematic evaluation of environmental noise. Up to now, various studies for finding the mutual relationship of this kind have been reported by many investigators. Almost all of these studies, however, were confined to only practical methods of directly applying the well-known linear regression analysis method to actually observed data and/or finding an approximate evaluation under the assumption of a standard Gaussian distribution form of level fluctuations. If one wants to find universally and objectively the more precise relationship among these evaluation indices, one cannot use the above practical method. From such an essential point of view, in this paper, a general theory for the mutual relationship among several type noise evaluation indices in close relation to two typical indices, \( L_\text{x} \) and \( L_\text{eq} \), has been proposed in a general form including the well-known simplified relationship based on a standard Gaussian distribution as a special case.

ESTIMATION THEORY OF \( L_\text{eq} \) EVALUATION INDEX BASED ON STATISTICAL INFORMATION ON LEVEL FLUCTUATION

Let us introduce the moment generating function, \( M_\text{eq}(\theta) \), with respect to the noise level fluctuation, \( L_\text{eq} \), which shows an arbitrary distribution form, as follows:

\[
M_\text{eq}(\theta) = \exp(\theta \eta + \theta^2/2)
\]

with

\[
\eta = x_\text{eq} \sigma_\text{eq} \sigma_\text{eq} \exp(-\theta^2/2),
\]

where \( \eta \) denotes the noise energy fluctuation and \( \eta \) denotes the reference noise energy usually taken as \( 10^{-12} \text{ (Watt/m}^2) \). Moreover, \( \sigma^2 \) denotes an averaging operation with respect to the random variable \( \eta \). The mathematical relationship between the arbitrary order cumulate, \( \kappa_n \), with respect to the moment generating function \( M_\text{eq}(\theta) \) is given by:

\[
M_\text{eq}(\theta) = \exp\left( \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} \theta^n \right)
\]

By replacing \( \eta \) to \( 1/\eta \) in Eq. (1) and (2), the following relationship can be easily obtained:

\[
\kappa_{n} = \frac{1}{n!} \exp\left( \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} \theta^n \right).
\]

Thus, a substitution of Eq. (3) into the definition of \( L_\text{eq} \) evaluation index yields a general series expansion type expression for estimating an \( L_\text{eq} \) evaluation index, as follows:

\[
L_\text{eq} = \sum_{n=1}^{\infty} \frac{\kappa_n}{n!} \eta^n
\]

where \( \mu = \kappa_1 \) is the mean value of \( L_\text{eq} \) and \( \sigma^2 = \kappa_2 \) is its variance. Especially when the noise level fluctuation shows a standard Gaussian distribution, Eq. (4) is reduced to a well-known usual estimation method:

\[
L_\text{eq} = L_\text{x} + 0.115 \sigma^2 = L_\text{x} + 0.115 \sigma^2,
\]

since the higher order cumulants \( \kappa_n (n=3,4,\ldots) \) for this case are equal to zero.

On the other hand, a general explicit expression of the cumulative noise level distribution function is expressed in the general form of a statistical Hermite series expansion expression [2], as follows:

\[
Q(L) = \frac{1}{\sigma^2} \exp[-(L-\mu)^2/2\sigma^2],
\]

where \( N(\mu, \sigma^2) \) denotes the well-known Gaussian distribution function:

\[
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp[-(x-\mu)^2/2\sigma^2] dx.
\]

and \( H_n(x) \) is the \( n \)th order Hermite polynomial. According to the definition of \( L_\text{x} \) sound level (i.e., \( 100\% \) percentage point of the level distribution), one can easily obtain the following relationship:

\[
1 - \frac{x}{100} = \frac{1}{\sigma^2} \exp[-(L-\mu)^2/2\sigma^2],
\]

where \( H_n(\cdot) \) is the \( n \)th order Hermite polynomial. Accordingly, the application of \( N \) kinds of specific \( L_\text{x} \) sound levels \( L_{x1}, L_{x2}, \ldots, L_{xn} \) to Eq. (6) yields the \( N \)-dimensional simultaneous equations for each \( \kappa_n (n=3,4,\ldots,N+2) \). This solution for \( N=2 \) is derived for practical use, as follows:

i) \( N=1 \) (by use of \( L_{x1} \)):

\[
L_\text{eq} = L_{x1} + 0.115 \sigma^2 = L_{x1} + 0.115 \sigma^2
\]

ii) \( N=2 \) (by use of \( L_{x1} \) and \( L_{x2} \)):

\[
L_\text{eq} = L_{x1} + 0.115 \sigma^2 = L_{x1} + 0.115 \sigma^2
\]

ESTIMATION THEORY OF \( L_\text{x} \) EVALUATION INDEX BASED ON STATISTICAL INFORMATION ON ENERGY FLUCTUATION

In this section, let us consider a unified estimation method of the arbitrary \( L_\text{x} \) sound level by use of the known values of \( L_\text{eq} \) and specific \( L_\text{x} \) evaluation indices. At the starting point of the analysis, as a general expression of a noise energy distribution function, a general form of a statistical Laguerre
expansion series can be employed [2]. The cumulative distribution function of the above expression is given after introducing a dimensionless variable, $\eta = (E/E_0)$, as follows:

$$Q(\eta) = \int_0^\eta P(t;m)dt + A_3 \frac{\eta^3}{3!} E_0^3 \exp(-\eta)F_c^2(\eta) + \cdots,$$

where $E_0$ and $\eta_0^2$ are the mean values of $E$ and its variance, respectively. Moreover, $L_0(m-1)(\cdot)$ denotes the associated Laguerre's polynomial. The distribution parameters, $m$ and $s$, can be rewritten by using an Leq evaluation index, as follows:

$$m = 10^{0.2L_0^P - 2.1\log_{10}E_0 \log_{10}E_0},$$
$$s = 10^{0.2L_0^P - 2.1\log_{10}E_0 \log_{10}E_0} \cdot$$

From Eq.(11) and the definition of $L_X$ sound levels, one can easily find the following relationship:

$$1 - \frac{x}{100} = \int_0^{L_0} P(t;m)dt + A_4 \frac{\eta_4^4}{4!} E_0^4 \exp(-\eta_4)F_c^2(\eta_4) + \cdots,$$

Thus, the application of $N$ kinds of specific $L_X$ sound levels $(L_0, L_0, \ldots, L_0)$ to Eq.(14) yields the $N$-dimensional simultaneous equations for $A_0 (n=1, 2, \ldots, N+2)$. That is, after estimating two distribution parameters, $m$ and $s$, by use of the known values of Leq and specific $L_X$, one can evaluate the arbitrary $L_X$ sound level by substituting the estimated value of $A_0$ into Eq.(11). At this time, the probabilistic relationship : $Q(L_X) = Q(\eta_0)$ is used.

The concrete functional form of estimating $A_0$ for $N=1$ or 2 is derived for practical use, as follows:

1) $N=1$ (by use of $L_0$):

$$A_3 = \frac{G(x)}{1 - \int_0^{\eta_0} P(t;m)dt}, \quad \eta_0 = \frac{E_0}{10} L_0/10,$$

2) $N=2$ (by use of $L_0$ and $L_0$):

$$A_4 = \frac{G(x)}{1 - \int_0^{\eta_0} P(t;m)dt}, \quad \eta_0 = \frac{E_0}{10} L_0/10,$$

Experimental Consideration

For confirming the validity of the proposed theory, the actually observed data of a road traffic noise has been used here. Table 1 shows the estimated results for the Leq evaluation index by use of the proposed method given by Eq.(9) or Eq.(10) and the usual method given by Eq.(5).

![Fig. 1](image) A comparison between the theoretically estimated curves by use of the proposed method and the experimentally sampled points for the cumulative noise level distribution. From these results, the estimated values are in good agreement with the experimental values.

<table>
<thead>
<tr>
<th>Observed value of Leq (dB)</th>
<th>Estimated values by use of Eq.(9) (dB)</th>
<th>Estimated values by use of Eq.(10) (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>73.7</td>
<td>75.9</td>
<td>74.6 (by L_0)</td>
</tr>
<tr>
<td>74.7</td>
<td>74.7 (by L_0)</td>
<td>74.4 (by L_0 and L_0)</td>
</tr>
<tr>
<td>74.3</td>
<td>74.3 (by L_0 and L_0)</td>
<td>74.3 (by L_0 and L_0)</td>
</tr>
</tbody>
</table>

REFERENCES

ETUDE DE PERCEPTION RELATIVE A LA DYNAMIQUE DU BRUIT DE LA CIRCULATION AUTOMOBILE ET A SON EVOLUTION AU TRAVERS D'UNE FAÇADE

B. Louvigné, J.-G. Migneron, M. Asselineau

Laboratoire d'acoustique et Centre de recherches en aménagement et en développement, 1624 pavillon F.-A. Savard, Université Laval, Ste-Foy, Québec, G1K 7P4

La dynamique des niveaux de bruit communautaire a fait l'objet d'une étude extensive dans la période 1965-75, suite notamment aux travaux londoniens du "Committee on the problem of noise" et à l'apparition de dispositifs simples pour l'analyse de la distribution statistique. Ces travaux ont culminé dans les années 70, avec les recherches parallèles de la B.R.S. en Grande-Bretagne et du C.S.T.B. en France, et la publication de la norme ISO 1996. Qu'on pense simplement à la floraison d'indices de bruit basés sur "le temps de passage" de la dynamique comme le TNT ("Transportation Noise Index") de Griffiths et Langdon ou bien aux savantes études de gêne de Aubée, Auzou et Kapin [1, 2, 3].

Paradoxalement, alors que les outils analytiques s'affirmèrent, on a vu depuis lors un net recul des paramètres statistiques de niveau de bruit au profit de l'emploi généralisé du seul niveau continu équivalent Leq, tant en ce qui concerne la métrologie, les normes et les législations, que la modélisation et les études d'impact [4, 5]. La recherche présentée ici tend à démontrer que la gêne ressentie dans les logements évolue de façon beaucoup plus délicate qu'il n'y paraît au premier abord [6].

Dispositif expérimental

Cette étude réalisée au Laboratoire d'acoustique de l'Université Laval fait suite à une analyse détaillée des caractéristiques physiques de l'isolement acoustique d'un échantillon de façade. Du côté sourcil, le dispositif expérimental retenu a permis de contrôler à la fois le niveau de pression, la composition spectrale, la dynamique et la directivité. Ces derniers paramètres ayant été ajustés par intensité-métrie, de façon à obtenir un champ acoustique normal [7].

Du côté récepteur, la petite chambre réverbérante du laboratoire a été traitée avec des matériaux absorbants (plafond suspendu et panneaux muraux), de façon à réduire son temps de réverberation aux enveloppes de 0,5 s; ceci à la fois pour permettre des mesures intensimétriques de dispersion et pour accommoder les personnes participant aux tests de perception (voir fig. 1). Le niveau de bruit de fond contrôlable a été obtenu à l'aide de quatre enceintes acoustiques balancées individuellement et installées dans le plafond suspendu, (échelle calibré conformément aux courbes NR).

Expériences à niveau équivalent constant

Les niveaux de six rubans échantillons de circulation routière (d'une demi-heure chacun) ont été ajustés à la reproduction de façon à obtenir un niveau continu équivalent Leq = 65 dBA, tel que relevé à 0,60 m de l'échantillon de fenêtre. Au cours des expériences les spectres émis ont été constamment visualisés sur un analyseur en temps réel et le comptage statistique des niveaux (notamment l'affichage du Leq) contrôlé à l'aide d'un analyseur statistique ("B & K" 4426). Il a été tout d'abord vérifié que la distribution statistique réelle des niveaux de bruit, relevée au cours des enregistrements en bordure des voies de circulation, puisse être reproduite intégralement devant l'échantillon de façade. Les six rubans retenus correspondaient à des tracés routiers compris entre 670 et 5972 v/h, avec des pourcentages de poids lourds compris entre 1,4% et 12,3%. Quant à la dynamique totale, elle s'étendait de 17,3 à 43,3 dBA, pour l'écart type de 15,8 à 35 avec les niveaux équivalents égaliées à 65 dBA).

Deux modes ont été simulées autour de cette dynamique dite "normale" d'un même échantillon de bruit routier, soit la "compression" et l"expansion", ces deux dernières variations de niveaux de bruit ayant été obtenues électromécaniquement (expanseur "Heath" AD-1700, et compresseur "Orban" 412A) (voir fig. 2). Pour tous les échantillons, les circuits d'amplification et de commutation ont été calibré de façon à assurer la constance du niveau Leq en façade.

Interprétation suivant les individus

L'incidence des différentes dynamiques a tout d'abord été écoutées, pour les six échantillons de bruit, par l'intermédiaire de six cas d'écoute dont les résultats ont été ensuite comparés aux mesures acoustiques; les sujets testés devant exprimer leurs évaluation des modes compressé et expansé en comparaison avec le ruban normal. Les bandes étaient écoutes les unes à la suite des autres, les sujets ont pu se repérer au mode normal de la bande précédente pour débuter l'évaluation d'une nouvelle bande. Il leur est demandé:

- de quantifier leur gêne par un chiffre allant de 1 à 7,
- et d'estimer, telles qu'elles peuvent les percevoir, les variations de vitesse, de proximité et de débit.

Le dépouillement des tests, on observe tout d'abord qu'une hiérarchie de la gêne s'établit pour chaque échantillon, que cette hiérarchie diffère suivant l'échantillon, et aussi qu'elle peut différer suivant les individus. Pour une même ouverture de fenêtre, les niveaux Leq étant maintenus constants en façade, les différences de gêne observées sont entièrement dépendantes des différences de dynamique du bruit. L'écoute type moyen des notes attribuées par les sujets pour trois positions de la fenêtre diminue de l' 0,7 puis 0,6, de la fenêtre fermée à la fenêtre complètement ouverte, plus la fenêtre est ouverte et moins les résultats de gêne sont donc dispersés. La configuration de la fenêtre fermée est la plus délicate à traiter, puisque la perception du bruit semble plus difficile et l'évaluation de la gêne plus subtile.

Les résultats du tableau 1 montrent que quelle que soit l'ouverture de la fenêtre, une partie des sujets
Tableau 1: Moyenne des réponses pour cinq individus en fonction de la dynamique.

<table>
<thead>
<tr>
<th>Fenêtre</th>
<th>fermée</th>
<th>entravée (15%)</th>
<th>ouverte (20%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>gêne moy.</td>
<td>écart</td>
<td>gêne moy.</td>
</tr>
<tr>
<td>Norm.</td>
<td>3.16</td>
<td>1.46</td>
<td>5.00</td>
</tr>
<tr>
<td>Comp.</td>
<td>4.33</td>
<td>2.13</td>
<td>6.00</td>
</tr>
<tr>
<td>2</td>
<td>2.83</td>
<td>1.34</td>
<td>4.33</td>
</tr>
<tr>
<td>Norm.</td>
<td>3.25</td>
<td>0.69</td>
<td>5.00</td>
</tr>
<tr>
<td>Comp.</td>
<td>4.50</td>
<td>1.38</td>
<td>6.66</td>
</tr>
<tr>
<td>3</td>
<td>3.66</td>
<td>0.74</td>
<td>5.16</td>
</tr>
<tr>
<td>Norm.</td>
<td>3.99</td>
<td>1.23</td>
<td>6.66</td>
</tr>
<tr>
<td>Comp.</td>
<td>4.85</td>
<td>1.16</td>
<td>6.00</td>
</tr>
<tr>
<td>4</td>
<td>3.66</td>
<td>0.74</td>
<td>5.00</td>
</tr>
<tr>
<td>Norm.</td>
<td>3.08</td>
<td>0.53</td>
<td>4.90</td>
</tr>
<tr>
<td>Comp.</td>
<td>2.90</td>
<td>0.53</td>
<td>4.40</td>
</tr>
<tr>
<td>5</td>
<td>4.00</td>
<td>0.73</td>
<td>4.33</td>
</tr>
<tr>
<td>Norm.</td>
<td>4.90</td>
<td>0.60</td>
<td>5.58</td>
</tr>
<tr>
<td>Comp.</td>
<td>2.90</td>
<td>1.53</td>
<td>4.00</td>
</tr>
</tbody>
</table>

Si deux tendances partagent la population dans la détermination et l'évaluation de la gêne, l'unanimité est faite quant à l'incidence de la variation de la dynamique sur le comportement psycho-perceptuel. Les notes attribuées à l'augmentation relative de vitesse, proximité et bruit, pour les modes compressé et expédié par rapport au mode normal, montrent que l'expansion du signal cause l'implication d'un rapprochement des véhicules mais surtout une augmentation de leur vitesse, alors que la compression du signal provoque une sensation nette d'augmentation du débit de circulation. Ce sont ces résultats qui ont été présentés, selon une évaluation en 5 points, dans le tableau 2.

Tableau 2: Impression psycho-perceptuelle en fonction de la modification de la dynamique.

<table>
<thead>
<tr>
<th>Fenêtre</th>
<th>fermée</th>
<th>entravée (15%)</th>
<th>ouverte (20%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode</td>
<td>indice de perc.</td>
<td>nb. cas</td>
<td>indice de perc.</td>
</tr>
<tr>
<td>Norm.</td>
<td>2.90</td>
<td>2/6</td>
<td>2.92</td>
</tr>
<tr>
<td>Comp.</td>
<td>3.25</td>
<td>4/6</td>
<td>3.30</td>
</tr>
<tr>
<td>2</td>
<td>3.07</td>
<td>2/6</td>
<td>3.20</td>
</tr>
<tr>
<td>Norm.</td>
<td>3.69</td>
<td>4/6</td>
<td>3.00</td>
</tr>
<tr>
<td>Comp.</td>
<td>3.37</td>
<td>5/6</td>
<td>2.69</td>
</tr>
<tr>
<td>3</td>
<td>2.76</td>
<td>2/6</td>
<td>2.78</td>
</tr>
</tbody>
</table>

Interprétation suivant les paramètres de bruit.

L'influence de la transmission au travers de la façade sur le dynamique a déjà été mentionnée [6], de même que son influence sur la composition spectrale [6, 7]. Les mesures préalables d'isolation dans différentes conditions d'ouverture montrent bien ce dernier effet; alors que les analyses statistiques des niveaux de bruit réalisées à l'intérieur du logement simulé montrent, dans la plupart des conditions, une diminution des écarts L10-L90 ou L10-L90. Il faut donc s'attendre, étant donné la dualité des perceptions mentionnées précédemment, à une modification possible de la classification des gênes fenêtre fermée, entravée ou ouverte.

Dans les conditions de fenêtre fermée, on peut trouver au tableau 3 l'exemple des rubans n° 4 et 5. Pour le premier de ces rubans (670 véhicules/h), le niveau des pointes et la forte dynamique déterminent la plus grande gêne, ceci malgré les 6 dB d'écart entre les niveaux Léq des différents modes. Alors que pour l'échantillon 5 (2034 véhicules/h), plus le niveau de bruit de fond est élevé et la dynamique réduite, plus la gêne est ressentie. De même à titre d'exemple, dans les conditions de fenêtre entravée (15%), on peut voir au tableau 4 les résultats correspondant au ruban

Tableau 3: Echantillons n° 4 et 5 fenêtre fermée.

<table>
<thead>
<tr>
<th>Mode</th>
<th>L10</th>
<th>L90</th>
<th>Leq</th>
<th>indice de gêne</th>
<th>écart o</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm.</td>
<td>29.0</td>
<td>37.4</td>
<td>30.2</td>
<td>5.41</td>
<td>1.09</td>
</tr>
<tr>
<td>Comp.</td>
<td>36.3</td>
<td>37.4</td>
<td>36.0</td>
<td>4.46</td>
<td>1.06</td>
</tr>
<tr>
<td>4</td>
<td>36.3</td>
<td>36.3</td>
<td>36.0</td>
<td>4.46</td>
<td>1.06</td>
</tr>
<tr>
<td>Comp.</td>
<td>36.3</td>
<td>37.4</td>
<td>36.0</td>
<td>4.46</td>
<td>1.06</td>
</tr>
</tbody>
</table>

n° 1 (2512 véhicules/h). La dynamique importante et les pointes déterminent la même gêne le plus faible (en mode expédié); alors que pour les autres modes, le bruit de fond devient dominant puisqu'il place en second le mode compressé.

Tableau 4: Echantillon n° 1 fenêtre entrouverte.

<table>
<thead>
<tr>
<th>Mode</th>
<th>L10</th>
<th>L90</th>
<th>Leq</th>
<th>indice de gêne</th>
<th>écart o</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm.</td>
<td>56.3</td>
<td>58.3</td>
<td>58.3</td>
<td>6.00</td>
<td>0.45</td>
</tr>
<tr>
<td>Comp.</td>
<td>53.3</td>
<td>58.3</td>
<td>58.3</td>
<td>6.00</td>
<td>0.45</td>
</tr>
<tr>
<td>4</td>
<td>50.3</td>
<td>58.3</td>
<td>58.3</td>
<td>6.00</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Conclusions

L'adoption d'une norme de bruit en niveau équivalent Leq en bordure d'une voie routière n'est donc pas toujours significative de la gêne réelle ressentie à l'intérieur d'un logement voisin. Les autres paramètres statistiques permettent de nuancer la prévision possible de la nuisance, en tenant compte du fait que pour la moyenne des répondants, lorsque la dynamique est élevée, la présence des pointes de bruit est la plus désagréable (indépendamment du débit de circulation), alors que lorsque cette dernière tend à diminuer (1100-1200 < 5 dB environ) c'est le niveau de bruit de fond Leq qui ordonne l'importance de la gêne; certains des répondants pouvant passer d'une sensation à une autre assez facilement, suivant les conditions de transparence de la façade et le niveau de bruit de fond du logement. Cette recherche subventionnée par le CSRC (Ottawa) et le ministère des Transports (Québec), doit se poursuivre, avec notamment l'introduction simulée d'un changement de pente dans l'isolation de la façade.

Références

A Unified Probability Expression for the Environmental Noise Fluctuating Only in a Finite Amplitude Region

Y. Fujita*, M. Ohta** and K. Hatakeyama**

*Faculty of Engineering, Hiroshima University, Shitami, Saio-Cho, Higashi-Hiroshima, 724 Japan

**Kure Technical College, Kure, 737 Japan

INTRODUCTION

In general, in the measurement of the actual stationary random noise, the amplitude fluctuation of the observed instantaneous value is limited within a finite range based on the mechanism of measurement equipment. Taking this limitation into consideration, the authors proposed already several kinds of series expansion type probability expressions related to Legendre's, Tchebyscheff's and Gegenbauer's polynomials by using the appropriate orthogonal function defined in the interval [-1, 1].

In this paper, a new unified expression by use of the Jacobi's polynomial defined in an interval [0, 1] is proposed in order to estimate the probability distribution function of the environmental noise. Furthermore, by choosing reasonably two distribution parameters of this Jacobi's polynomial type probability density expression, it is found that the proposed expression coincides with Gegenbauer's polynomial series expansion type probability density expression and contains the well-known Hermite and Laguerre series expansion type probability expression, in the limiting case when a distribution parameter tends to infinity.

Finally, the justification and effectiveness of the proposed probability expression will be confirmed by comparing this theoretical distribution curve with experimentally sampled values of a road traffic noise.

THEORETICAL CONSIDERATION

Let us find out a general form of the orthogonal series expansion type probability density function matched to the limitation of fluctuation range based on the measurement mechanism. First, without neglecting the generality of analysis, we consider an ideal case where the stationary random noise fluctuates within the amplitude interval [0, 1]. Then, the probability density function is expanded in the following orthogonal series:

\[
P(x) = \sum_{n=0}^{\infty} a_n \phi_n(x),
\]

where

\[
a_n = \frac{\int_{0}^{1} \phi_n(x) p(x) \, dx}{\int_{0}^{1} \phi_m(x) \phi_n(x) \, dx}
\]

and \( p(x) \) is satisfied the normalized condition:

\[
\int_{0}^{1} p(x) \, dx = 1.
\]

Here, \( \{ \phi_n(x) \} \) forms a complete set of orthonormal functions with a weighting function \( p(x) \):

\[
\int_{0}^{1} \phi_m(x) \phi_n(x) \, dx = \delta_{mn}.
\]

As an orthogonal function defined within the amplitude interval [0, 1], Jacobi's polynomial \( G_n(x) \) can be reasonably chosen:

\[
G_n(a, \alpha; x) = \sum_{i=1}^{n+1} \gamma_i (\alpha + i) \Gamma(\gamma + i) x^i,
\]

where

\[
a_n = \frac{\int_{0}^{1} \phi_n(x) G_n(a, \alpha; x) \, dx}{\int_{0}^{1} \phi_m(x) G_n(a, \alpha; x) \, dx}.
\]

Of course, Jacobi's polynomial satisfies the following orthogonal relation:

\[
\int_{0}^{1} G_m(a, \alpha; x) G_n(a, \alpha; x) x^\gamma (1-x)^{\alpha-\gamma} \, dx = \frac{\delta_{mn}}{\int_{0}^{1} \phi_m(x) G_n(a, \alpha; x) \, dx}. (6)
\]

By the well-known definition of beta function:

\[
\int_{0}^{1} x^{\gamma-1} (1-x)^{\alpha-\gamma} \, dx = \frac{\Gamma(\gamma) \Gamma(\alpha)}{\Gamma(\gamma + \alpha)}.
\]

Hence, \( p(x) \) in eq. (5) is concretely given as follows:

\[
p(x) = \frac{1}{B(\gamma, \alpha+\gamma)} x^{\gamma-1} (1-x)^{\alpha-\gamma}. (7)
\]

As is well known, \( B(p, q) \) can be directly expressed by using a gamma function \( \Gamma \):

\[
B(p, q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p+q)}.
\]

Using eq. (8), eq. (6) is easily rewritten as the following form:

\[
\int_{0}^{1} [x^{\gamma-1} (1-x)^{\alpha-\gamma}] \cdot \frac{\phi_m(x) G_n(a, \alpha; x)}{\phi_n(x) G_m(a, \alpha; x) \Gamma(\gamma) \Gamma(\alpha) \Gamma(\gamma + \alpha)} \, dx = \frac{\delta_{mn}}{\int_{0}^{1} \phi_m(x) G_n(a, \alpha; x) \, dx}.
\]

Comparing eq. (4) with eq. (10), the orthonormal function \( \phi_n(x) \) can be concretely expressed as follows:

\[
\phi_n(x) \Gamma(\alpha+\gamma) \Gamma(\gamma+\alpha) \Gamma(\alpha+\gamma+n) \Gamma(\gamma+\alpha+n) \frac{\phi_n(x) G_n(a, \alpha; x)}{\phi_n(x) G_n(a, \alpha; x) \Gamma(\gamma) \Gamma(\alpha) \Gamma(\gamma + \alpha)}.
\]

Also, using eq. (2), an expansion coefficient \( a_n \) can be evaluated:

\[
a_n = \frac{\int_{0}^{1} \phi_n(x) G_n(a, \alpha; x) \, dx}{\int_{0}^{1} \phi_n(x) G_n(a, \alpha; x) \, dx}.
\]

Thus, the probability density function of stationary random noise fluctuating only in the interval [0, 1] can be explicitly expressed in the orthogonal series expansion form:

\[
P(x) = \sum_{n=0}^{\infty} a_n \sum_{i=1}^{n+1} \gamma_i (\alpha + i) \Gamma(\gamma + i) x^i.
\]

That is, by using the measure-preserving transformation of probability, the following equation can be directly derived:

\[
P(x) = \frac{(X-a) Y^{\gamma-1} Y^{\alpha-\gamma}}{B(a, \alpha+\gamma) (b-a) b \phi_m(a, \alpha+\gamma)} \frac{1}{\phi_n(a, \alpha+\gamma)} \frac{\phi_n(x) G_n(a, \alpha; x)}{\phi_n(x) G_n(a, \alpha; x) \Gamma(\gamma) \Gamma(\alpha) \Gamma(\gamma + \alpha)}. (11)
\]

Using eq. (5), two expansion coefficient \( A_1 \) and \( A_2 \) can be given as follows:

\[
A_1 = \frac{\gamma+\alpha+1}{\gamma+\alpha+1} \frac{(X-a)}{b-a} - \frac{a+1}{\gamma+\alpha+1} \frac{a+1}{b-a}, (12)
\]

\[
A_2 = \frac{A_1}{(\gamma+\alpha+1)} \phi_{n=1}(a, \alpha+\gamma) \phi_{n=1}(a, \alpha+\gamma).
\]

The two distribution parameters \( a, b \) in eq. (16) are determined to satisfy the following relation:

\[
A_1 = 0, A_2 = 0.
\]
That is, by using two lower order statistical moments of mean $\mu$ and variance $\sigma^2$ of noise fluctuation, we easily have

$$\sigma^2 = \frac{(u-a)(b-u)}{u^2} - 2,$$

$$\gamma = \frac{(u-a)(b-u)}{u^2} - 1.$$  

(20)

Here, by using eq.(14), let us consider again eq.(13) derived in the universal expression with a non-dimensional variable $x/(\langle X^2 \rangle / (b-a))$. The probability density function $P(x)$ can be expressed in a series expansion form consisting of successive derivatives of beta function type p.d.f. with different parameters. Now, by introducing the definiton:

$$P_0(x;\alpha+2n,\gamma+n) = \frac{\alpha^{\gamma+n-1}(1-x)^{\alpha-n} \gamma^n}{B(\gamma+n,\alpha-\gamma+n+1)}$$  

(22)

differentiating this $P_0()$ n times, the n-order successive derivative can be found as follows:

$$P_0(n)(x;\alpha+2n,\gamma+n) = \frac{\partial^n}{\partial x^n} P_0(x;\alpha+2n,\gamma+n)$$

$$= \frac{1}{B(\gamma+n,\alpha-\gamma+n+1)} \frac{\partial^n}{\partial x^n} \left( \frac{\gamma^{\gamma+n}(1-x)^{\alpha-n}}{B(\gamma+n,\alpha-\gamma+n+1)} \right).$$

(23)

From the definition of Jacobi's polynomial, eq.(5), the following equation can be easily obtained:

$$\frac{d^n}{dx^n} (1-x)^{\alpha-n} \gamma^n = \frac{\Gamma(n+\gamma)}{\Gamma(n)} (1-x)^{\alpha-n} \gamma^n.$$  

(24)

Then, the following expression can be obtained:

$$P_0(x;\alpha,\gamma) = \frac{\Gamma(n+\gamma)(\alpha+n+1)}{\Gamma(n+\alpha+1)} \frac{\partial^n}{\partial x^n} \left( \frac{(1-x)^{\alpha-n} \gamma^n}{B(\gamma+n,\alpha-\gamma+n+1)} \right).$$

(25)

Thus, by substituting eq.(25) into eq.(23), the p.d.f. of $x$, $P(x)$, can be given in the following series expansion form:

$$P(x) = P_0(x;\alpha,\gamma) + \sum_{n=1}^{\infty} B_n P_0(n)(x;\alpha+2n,\gamma+n),$$

(26)

$$B_0 = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+2n+1)\Gamma(\alpha+1)}.$$  

(27)

In the actual engineering field, the cumulative distribution function is more important than the probability density function. The cumulative distribution function is directly derived as follows:

$$Q(x) = \int_0^x P_0(y;\alpha,\gamma) dy + \sum_{n=1}^{\infty} A_n \frac{\gamma^n}{\Gamma(n+2\alpha)}.$$

(28)

Here, as three special cases of eq.(16), the following expression of p.d.f. can be derived:

1. Statistical Gegenbauer expansion type p.d.f. expression ($\alpha=2, \gamma=1, \gamma=1/2$):

$$P(x) = \frac{1}{B(\alpha+1/2,1/2)(b-a)/2} \left[ 1 - \frac{X-(a+b)/2}{(a-b)/2} \right]^{\gamma-1/2}$$

$$\times \left[ 1 + \sum_{n=1}^{\infty} A_n C_n^\alpha \left( \frac{X-(a+b)/2}{(b-a)/2} \right)^2 \right]^{\gamma-1},$$

(29)

2. Statistical Hermite expansion type p.d.f. expression ($\alpha=2, \gamma=1, \gamma=1/2, \gamma=1/2$):

$$P(x) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \left[ 1 + \sum_{n=0}^{\infty} A_n H_n(x-\mu) \left( \frac{x-\mu}{\sigma} \right) \right],$$

(31)

$$A_n = \frac{\gamma^n}{n!} \left( \frac{x-\mu}{\sigma} \right)^n,$$

(32)

where $H_n()$ means a Hermite polynomial of order $n$.

3. Statistical Laguerre expansion type p.d.f. expression ($\gamma=n, \gamma=n, \gamma=n, \gamma=n$):

$$P(x) = \frac{1}{B(\alpha+n, \gamma+n)} \exp \left[ -\frac{X}{\sigma} \right] \left[ 1 + \sum_{n=0}^{\infty} A_n L_n(m-1) \left( \frac{x}{\sigma} \right) \right],$$

(33)

$$A_n = \frac{\gamma^n}{n!} \left( \frac{x}{\sigma} \right)^n,$$

(34)

where $L_n(m)()$ means a Laguerre polynomial of order $n$ with parameter $m$.

Owing to the page limitation, its detailed discussion is omitted here.

**EXPERIMENTAL CONSIDERATION**

The validity of our proposed method has been experimentally confirmed by applying it to the actual road traffic noise in Miyoshi City. Figure 1 shows a comparison between experimentally sampled points and theoretically evaluated curves in a case with a fluctuating range [60,90] dB(A). Theoretical probability distribution curve has been calculated by using the third approximation of the statistical Jacobi series type expansion and the statistical Hermite series type expansion. The approximated curve by the statistical Jacobi series type expansion vibrates largely and is not shown. The experimental result clearly shows the usefulness of the proposed expansion.

**CONCLUSION**

In this paper, the probability expression of statistical Jacobi's series expansion type has been newly proposed in a new form matched to its finite fluctuation range based on the measurement mechanism. The validity and effectiveness of the proposed theory has been experimentally confirmed by applying it to the actual road traffic noise.

**ACKNOWLEDGEMENT**

We would like to express our cordial thanks to Mr.K.Murakami, Mr.J.Nakamura and Dr.H.Yamada for their helpful assistance.

**REFERENCES**

1. S.Moriguchi and et al., Tables of Function with Formulas, p.87-96 (1977).
NOISE PREDICTION MODEL IN HIGH INTENSITY HIGHWAYS

E. Gaja Diaz
F. Belmar Ibáñez
M. Estelles Belenguer
P. Solana Quiros

Departamento de Física. Universidad Politécnica de Valencia. España.

INTRODUCTION
The following conclusions were drawn out from a poll recently carried out by the Physics Department.
* 81.4% of the inquired say that the street where they live is very or quite noisy.
* 80.3% of the inquired people say that the greatest noise comes from traffic.
Let us such data as a basis for the necessity of developing a mathematical model able to predict the noise caused by urban traffic noise. Due to the difficulty and extent of this topic, we decided to develop a model under certain conditions:
- High traffic intensity highways.
- Highways with straight and horizontal stretches.
- Traffic flow without stop-points.
- Noise propagation in urban environments.

NOISE SOURCE FEATURES
Before developing the theoretical model, the first step to follow was to analyze the noise source, that is, to study the motorcar as the main traffic noise source. This study was centered in 3 basic features: a) pressure level measurements, depending on speed b) type of acoustic source caused by a motorcar, c) frequency analysis of the emitted noise.

a) In relation to this point some samples were taken for different vehicles at different speed past two sound level meter placed at 4 and 8 meter from the basis of the perpendicular.
- Light vehicles at 4 meter:
  \[ L_p = 40.9 + 15 \log v \]
- Heavy vehicles at 4 meter:
  \[ L_p = 44.5 + 20 \log v \]

b) Two studies were carried out to check whether vehicles assimilable to point sources. First, records of different vehicles at different speed were taken with a magnetic recorder, comparing this track with the one it should have in case of emitting as a point source. (See F.1). Second, analysis of the difference in the \( L_p \) obtained for 4 and 8 meter at different speeds. (See F.2). From these two experiences we have concluded that a single vehicle can be assimilated to an ideal point-source.

c) The frequency analysis was carried out with bands of thirds of octave. We concluded that the bands with most energetic content are always found under 1000 Hz.

INTERACTING FACTORS
A detailed study of all factors was done, coming to the conclusion that the most significant factors are: vehicle flow, traffic speed, number of lane, traffic composition, wave propagation, facade and cannon effect.

DEVELOPMENT OF THE MODEL
From the equation obtained for \( L_p \), the steps we followed are:
1) Establishment of equivalent level for a single vehicle. Having into account that pressure level in function of time and that the vehicle is a point-source, for a distance \( d' \) is:

\[ L_p(t) = L_{p\text{Max}} - 10 \log \left(1 + \left(\frac{v}{d'}\right)^2\right) \]  \hspace{1cm} (1)

We can obtain the equivalent level of a single vehicle by substituting the equation:

\[ L_p(t) = 10 \log \frac{1}{T} - 10L_{10}\frac{10}{10} \int dt \] \hspace{1cm} (2)

then the equation will be:

\[ L_p = L_{p\text{Max}} + 10 \log \frac{d'}{d} \left(\text{arctg} \frac{v}{2d}\right) \] \hspace{1cm} (3)

2) Establishment of equivalent level for a given traffic flow. The noise source was considered as a linear source consisting of several points sources corresponding to single vehicles. Let us suppose that vehicles belong to two types: light and heavy vehicles and that every vehicle belonging to one type is similar to an ideal average vehicle with noise level depending on the speed previously measured. With this statement and having into account that for this kind of sources:

\[ L_e = 10 \log \frac{1}{T} - 10L_{10\text{t}} \] \hspace{1cm} (4)

and that in 1 hour time \( N \) vehicles flow, we have the equation:

\[ L_e = L_{p\text{Max}} + 10 \log \frac{d'}{d} \left(\frac{\log N}{10}\right) \] \hspace{1cm} (5)

Therefore there are two factors to be considered for \( L_e \): light vehicle contribution and heavy vehicle contribution with their different \( L_{p\text{Max}} \). Substituting in equation (5), it we will have:

Light: \[ L_e = 10 \log \frac{d'}{d} + 5.1 \log v_1 + 15.9 \] \hspace{1cm} (6)

Heavy: \[ L_e = 10 \log \frac{d'}{d} + 10 \log v_1 + 19.5 \] \hspace{1cm} (7)

3) Establishment of equivalent level in function of distance. These equation (6),(7) refer to a distance \( d' = 8 \) meter as the \( L_{p\text{Max}} \) was calculated for a such a distance. Therefore in order to use these equation for other distance and having into account that the source is a line, we must add the term \( 10 \log d'/d \) to the previous equations. Then we would have:

Light: \[ L_e = 10 \log \frac{d'}{d} + 5.1 \log v_1 - 10 \log d + 34 \] \hspace{1cm} (8)

Heavy: \[ L_e = 10 \log \frac{d'}{d} + 10 \log v_1 - 10 \log d + 37.5 \] \hspace{1cm} (9)

And the equation for total equivalent level will be:

\[ L_e = 10 \log \left(10^{-10\text{eq}}/10 + 10^{-10\text{eq}}/10\right) \] \hspace{1cm} (10)

4) Equivalent lane calculation. If we observe equation (8),(9) we can note there is an only distance \( d \), which isn't actually true as in city road there are several lane with the corresponding variety of distance-observer-traffic lane. Thus, we elaborated a term called "equivalent lane" which is a fictitious lane where all traffic must flow in such a way that the \( L_e \) caused by traffic on a certain observer the same as the \( L_e \) produced by traffic distribution among all the lanes. 3 In F.3 distance under certain conditions can be determined.

5) Introduction of correcting factors. The last step was to determine the correction factors. We previously observed if the model was suitable for city-road without building in the nearby. Once it was checked we deduced that the streets with facade to one and/or both sides, had different values due to facade and cannon effects. First we calculated the facade effect in streets with building at one side only, obtaining an average correction of 2.5 dB(A). Second we calculated the cannon effect in streets with buildings to both sides, obtaining a relationship between noise level increase and rate between building height and street width (J). (See F.4)
MODEL TESTING

This model was compared with real measurements from traffic (light and heavy vehicle), speed (light and heavy vehicles) and noise level in 32 streets of Valencia where the required conditions were accomplished. In each street samples were taken for 12 hours, for 8 a.m. to 8 p.m. This model was submitted to a correlation analysis with the experimental results, and we obtained a correlation of 80%. It was also submitted to a normality test with a readability threshold of 95%.

PRACTICAL APPLICATION OF THE MODEL

Once the prediction model has been obtained and checked, the last step to follow is to improve the model as we have two equations which depend on many variables. Therefore a study of the light and heavy vehicles speed, we coming to the conclusion that:
- light vehicles: 65% of speed between 30 and 45 Km/h
- heavy vehicles: 50% of speed between 25 and 35 Km/h

If we consider 37.5 Km/h the most representative speed for light vehicles and 30 Km/h for heavy vehicles, then heavy traffic will be $Q = Q_p$, where $Q$ is total traffic and $p$ the heavy vehicles percentage, the equation resulting from summing up light and heavy equivalent level with such conditions will be:

$$L_{eq} = 10 \log Q - 10 \log d + 10 \log(1 + 16.5p) + 42 \ (11)$$

From equation (11) a figures has been drawn where different variable for $L_{eq}$ calculation in function of $Q, p$ or $d$ are shown. By adding the correction factors we obtain the $L_{eq}$ more easily. (See F-5

REFERENCES

4) MCHRP. "Highway noise. A design guide for highway engineers." Report 117. 1971
5) HER MAJESTY'S STATIONERY OFFICE. "A guide to measurement and prediction of the equivalent continuous sound level." London. 1978
Développement de modèles informatiques pour l'étude d'impact des autoroutes et le calcul des écrans acoustiques

P. Jacques, A. Esteve, J.-G. Migneron, M. Asselinneau

Laboratoire d'acoustique et Centre de recherches en aménagement et en développement. 1624 pavillon F.-A. Savard, Université Laval, Québec, G1K 7P4

Depuis les années 1960 on a clarifié le problème théorique de l'impact des sources linéaires en obtenant notamment les équations de référence:

\[ L_{50} = -20 \log b + C_{50} \text{ (en bordure des voies)} \]

\[ L_{50} = 10 \log Q - 10 \log d + C_{50} \text{ (à distance d)} \]

Leq = Leq - 10 \log d - 10 \log Q - 10 \log d + C_{leq}

avec la relation pratique \( Q = KV = V/b \) \((Q, K, V \text{ et } b, \text{ étant respectivement le débit horaire, la concentration de véhicules, la vitesse et l'espacement moyen dans un système d'unités cohérentes})\) [1, 2]. Ces analyses sont à l'origine de nombreux modèles plus ou moins théoriques ou pratiques, issus principalement de la régression entre le niveau de bruit mesuré au bord des voies et le logarithme du débit [1, 3, 4]. Ainsi, dans la région de Québec nous avons obtenu, en bordure des voies, les équations empiriques:

\[ L_{50} = 17 \log Q + 14.8 \]

\[ \text{Leq} = 11.3 \log Q + 37.8 \]

Néanmoins l'application d'un modèle simple à la réalité de circulation urbaine pose de nombreux problèmes, à savoir tout d'abord ou localiser l'isoéphémère de référence (avant tout calcul de propagation) [4] et ensuite, quelles sont les influences des paramètres complémentaires tels que la vitesse, l'état de la chaussée, la pente ou le pourcentage de poids lourds. Pour pallier notamment à l'effet du pourcentage de véhicules lourds (qui remonte le niveau des points et le Leq), une tâche de ces dernières années a consisté à définir un débit fictif avec un coefficient d'équivalence des véhicules lourds suivant la relation \( Q_{eq} = Q_{veh} \times \frac{1}{1 + V_{veh}/V_{veh}} \) [2, 5]; en outre, on a distingué séparément l'impact des différentes catégories de véhicules [4]. En ce qui concerne la vitesse, plusieurs approches ont été proposées (par exemple, de 20 à 30 \( \log V \)) [3, 4].

De notre côté, nous avons cherché à obtenir une expression de la relations de correction proposé en 1976 par la HRS, qui tient compte simultanément du pourcentage de véhicules lourds et de la vitesse [5]. En outre, la hauteur de la source doit pouvoir être ajustée, en fonction de la vitesse des véhicules et du pourcentage de poids lourds, de manière à prendre en compte la répartition de l'émission du bruit par les pneumatiques ou les échappements [5]. Plus élargément, il est possible d'accorder à chaque véhicule une puissance acoustique appropriée et de calculer un niveau de puissance linéaire par tonne de voie, suivant le débit, les conditions de circulation et le pourcentage des véhicules lourds [7]. Enfin, une dernière tendance a consisté à introduire le calcul simultané de coûts de construction des écrans de protection [8].

Modélisation de la propagation et des effets d'écrans

La modélisation de la propagation peut s'effectuer de deux manières différentes, suivant le modèle de source considéré. Dans la première approche, la voie de circulation est assimilée à une source linéaire (ou à des tronçons de droite vus sous des angles donnés), ce qui conduit à écrire le niveau de bruit résultant à la distance \( d \) de cette voie sous la forme:

\[ L = L_{eq} - 10 \log d - A_{eq} (d) + 10 \log \ell/180 \]

où \( L_{eq} \) est le niveau de référence de la source, et \( A_{eq} (d) \) un terme prenant en compte l'effet de sol. Dans la seconde approche, la voie de circulation est représentée par un ensemble de sources élémentaires, le niveau de bruit résultant, pour chacune des sources, étant obtenu à l'aide de la relation:

\[ L_{eq} = 10 \log \ell - 8 \log d_{i} - A_{eq} (d_{i}) \]

dans laquelle \( L_{eq} \) est le niveau de puissance acoustique moyen et \( d_{i} \) la distance du point d'écoute.

Dans les deux cas, vient donc s'ajouter un terme d'effet de sol, qui dépend de l'impédance moyenne du sol, de l'angle des talus bordant l'autoroute, de la présence de constructions ou d'écars végétaux, des conditions météorologiques, etc. De récentes études sur la propagation du bruit à grande distance nous ont confirmé la complexité de ce dernier terme [9]. Il est de cette façon particulièrement difficile à modéliser, les approches proposées étant souvent trop simplistes, ou bien au contraire trop complexes pour l'usage informatique. Cependant, une approche initialement proposée par la B.R.S. et reprise depuis par d'autres auteurs [6, 1], propose une expression de la forme:

\[ A_{eq} (d) = a \log b \left[ \frac{b}{c_{eq} (d_c + c)} \right] \]

Avec ce modèle, l'atténuation n'apparaît qu'au-dessous d'une droite d'équation \( h_{eq} = 1/b \) (eq. c). Pour obtenir un résultat correspondant à la situation réelle, les coefficients \( a, b, c \) doivent être ajustés, ce qui revient notamment à déplacer le point d'origine de la droite limite, pour tenir compte de l'effet de la topographie (voir fig. 1).

En ce qui concerne la modélisation des effets d'écrans, les diverses solutions proposées font toutes appel au paramètre de différence de parcours acoustique, mais elles se compliquent considérablement dès que le nombre d'écars augmente. En outre, ces solutions ne tiennent pas compte du changement de fréquence provoqué par l'effet du filtre de diffusion (en pratique, la longueur d'onde est prise comme une constante \( 0.5 \lambda m \)). La formule tirée de la théorie de Millau, qui donne pour l'écran droits \( 10 \log (0.5 \lambda / \ell) \) est trop optimiste quant à la réduction que l'on peut encompter d'un écran. Pour cette raison, on peut lui préférer la formule pratique mise au point par les chercheurs du C.N.R. [3, 5]:

\[ A_{eq} = 7.7 \log \ell + 13.7 \]

Modèle basé sur des plans de propagation acoustique

Le premier modèle développé au C.R.A.D. de l'Université Laval opère sur ordinateur "Hewlett-Packard" de la série 200, il permet de calculer des paramètres tels que \( L_{99\%}, L_{50\%}, \text{Leq ou Leq} \) (24 h); il intègre notamment les équations proposées par la S.C.H.L. à partir des travaux du C.N.R. [3]. Des paramètres de correction sont applicables à la vitesse, au pourcentage de poids lourds, à l'état de la chaussée, ainsi qu'à la hauteur et à la position du point-source équivalent. Les calculs d'écran et de propagation sont d'abord basés sur une bonne mise en coordination des principaux paramètres géométriques, tels que bordures des voies, pieds et sommets des écrans, bords des talus, topographie générale du terrain, localisation des fenêtres exposées, etc.

En ce qui concerne la propagation et l'effet de sol, le modèle retient le principe de la droite limite proposée par la B.R.S., cependant la solution est variable suivant la topographie et le profil en travers de l'autoroute. D'autre part, cette équation se trouve complétée dans le cas d'autoroutes en remblais ou en viaduc, afin de calculer l'atténuation pour les niveaux topographiques situés sous les voies (voir fig. 2). En ce qui concerne les effets d'écran, les procédures sont légèrement simplifiées de manière à pouvoir identifier automatiquement et calculer
l'effet des écrans simultanés pour une même voie de circulation.

Tous les paramètres principaux du modèle sont accessibles et modifiables, ce qui permet d'ajuster les calculs de niveaux source et de propagation à la réalité d'un site, une fois rentrées les coordonnées des points de mesure réels sur le plan de propagation et étudié. Ce premier modèle n'aborde cependant pas systématiquement la cartographie des niveaux résultant de la simulation. Par contre, il est toujours possible de relier les points de même niveau sonore des différents plans et ainsi obtenir un profil simplifié d'impact de l'autoroute.

Modèle général décomposant toutes les voies en tronçons simples, droits ou courbés

Étant donné la complexité relative de certains échangeurs et les fluctuations topographiques des voies principales ou des bretelles de raccordement, nous avons été amenés à prévoir un nouveau logiciel: "AUTO II". Il a été développé en rapportant les données du programme TRM-PC, dont les premiers développements ont été écrits en "BASIC Microsoft" version 3.0, mais la version finale est en "Turbo-Pascal", de façon à utiliser au maximum les avantages d'un langage compilé et du processeur mathématique 8087.

Ce programme "AUTO II", permet le calcul du niveau de bruit dû à des voies routières en divers points de l'espace, quelle que soit la position géométrique et la forme des voies considérées. Sur tout tronçon de voie simple, le bruit est considéré comme uniformément distribué, avec un niveau de puissance par unité de longueur établi à partir de puissances de référence, tant pour les véhicules légers que pour les poids lourds (en pratique les voies sont découpées en tronçons élémentaires de 10 m). Les valeurs des puissances acoustiques de référence, fonction de la vitesse, ont été obtenues à partir de la littérature et des résultats antérieurs du CRAD, elles sont fixées de la manière suivante:

- véhicules légers: de 0.10 à 0.66 W, entre 50 et 100 km/h,
- et véhicules lourds: de 0.30 à 1.100 W, également entre 50 et 100 km/h [7].

L'ensemble de ces voies routières est soumis à sections simples fixées par la topographie et la forme des tracés existants ou projetés. Il peut s'agir de sections rectilignes (définition par 2 points), ou de courbes (définition par 3 points). La hauteur de la source varie entre 0.5 et 2.5 m pour des vitesses de 100 à 50 km/h et des pourcentages de poids lourds compris entre 0.5 et 20% [5]. Tout comme notre logiciel précédent, "AUTO II" considère trois situations principales, soit le cas d'un terrain plat ou asimilé, le cas où l'autoroute est surélevée par rapport au terrain naturel et le cas où elle est en contreforts. D'autre part, le programme identifie tous les effets d'écarts, naturels ou dus à des dispositifs de réduction de bruit, et a recours à un programme de calcul considéré; ensuite ils les compile pour chacun des segments de voie. Finalement, l'impact acoustique de l'ensemble des voies routières est la somme des niveaux de bruit résultant des différents segments et tronçons considérés, tant sur les voies routières que sur les secteurs urbains qui ont été représentés.

Conclusions

Le premier modèle est séduisant en ce qui concerne le temps de calcul, d'autre part il permet d'obtenir différents paramètres, tels L50X ou L99%. Cependant, son application est limitée à l'analyse dans un plan de propagation, pris en général perpendiculaire à la ligne source. De plus, la voie de circulation doit pouvoir être assimilée à des segments de droites suffisamment longs.

Le second modèle, complétement en trois dimensions, permet aisément d'effectuer une cartographie automatique des isoniveaux correspondant aux niveaux de bruit équivalents calculés; tout en tenant compte, si nécessaire, de sources extérieures (voies de chemin de fer ou industrie), ainsi que des diverses particularités des voies de circulation étudiées. Le temps de calcul, dans ces conditions, est passablement long, mais nullement prohibitif.

Références


FIGURE 1: Conditions topographiques et effet de sol.

FIGURE 2: Sortie graphique fournie par le programme.
FILTER DESIGN OF ACTIVE SILENCER
FOR AIR-CONDITIONING DUCTS

T. Emokida, H. Hamada, T. Miura
M. Takahashi, T. Kuribayashi, K. Asami

(1) Tokyo Denki University
2-2, Kanda-Nishiki-cho, Chiyoda-ku,
Tokyo, Japan

(2) Hitachi Plant Engineering & Construction co., Ltd. Research Laboratory
537, Kami-Hongo, Matsudo-shi, Chiba-ken, Japan

1. INTRODUCTION

Our research is to design and realize an active silencer for air-flow noises in air-conditioning ducts. We proposed the Dual Sensing Microphones (DSM) system for actively silencing air-flow noises in ducts [1]. In this paper, we describe the principle, the configuration and the filter design including adaptive algorithms.

2. PRINCIPLE

In the typical monopole system as in Fig.1 (a), the anti-phase output signal (-Y) from the control system is synthesized from the given input signal (X) and the transfer function (H) in the ducts. On the other hands, in the new arrangement monopole system as in Fig.1 (b) the output signal (-Y) can be derived from the given signal difference (X-Y) and H, namely,

\[ -Y = \frac{1}{1-H} (X-Y) \]  

therefore if H is not equal to 1, -Y is expressed by eq. (1).

![Diagram of Typical Monopole System](image1)

![Diagram of New Arrangement Monopole System](image2)

Fig.1 Simple Models of Active Silencers

In general, -Y propagates to the both side microphones of the loudspeaker. The feedback sounds are incorporated in Fig.2. In the typical monopole system (Fig.2 (a)), the feedback sounds propagate to the sensing microphone in the upstream, where H1 denotes the transfer function from the output of the control system to the sensing microphone. In the new arrangement monopole system (Fig.2 (b)), the feedback sounds propagate to the two sensing microphones in the upstream and downstream, where transfer function H2 is defined in the same way as H1 besides the way to the downstream. While in Fig.2 (a), the feedback sounds may cause unstable state, the new arrangement monopole system in Fig.2 (b) proves to be stable if H1 is equal to H2. Because the feedback sound signals picked up two sensing microphones are mixed together and negated in the input of the control system. That is to say, the new arrangement monopole system could be stable if the loudspeaker is located in the center of the two sensing microphones.

3. THE DUAL SENSING MICROPHONES SYSTEM

The DSM system includes two sensing microphones (M1, M2) and a loudspeaker (S) which is located in the center of M1 and M2 to negate feedback sounds from S to the each microphone as in Fig.3 [1], where H0 denotes the transfer function of the cancellation filter.

Fig.4 [1] shows the block diagram of the DSM system. All the summations of the acoustic and electric signal are described electrically. Hr and Ht include all the acoustic and electrical properties from the output of the digital filter to M1 and M2 respectively. If Ht is equal to Hr, the feedback sounds is completely negated in the input signal of the digital filter. In practice, it is not equal completely because of differences of physical condition from S to M1 and M2. Fig.5 shows the measured closed loop gain (CLG) of the DSM system as in Fig.3. He(Hr+Ht) denotes the CLG of the DSM system, while HeHr may be formed the CLG of the typical monopole system. It is possible to allow the amplitude of CLG in the DSM system less than 1 (0dB), say stable, for the practical loudspeaker-microphones arrangements.
4. SYSTEM DESIGN

Let suppose that all the system components including a loudspeaker, two sensing microphones and acoustic properties in ducts are linear time-invariant.

It is considered to realize $H_e$ using a FIR digital filter. Well known filter design algorithm in the time domain, Least Square Algorithm (LSA) has been applied. Approximated inverse filter $h$ is represented by LSA as follow,

$$h = (X^T X)^{-1} X^T Y \quad (2)$$

where $X^T X$ is a auto-correlation matrix of the input time series, and is symmetric and Toeplitz. While $X^T Y$ is a cross-correlation matrix of the input time series to the output time series. Therefore we are able to use FFT and Levinson algorithm [2] to calculate $h$.

5. ADAPTIVE CONTROL

We supposed that all the electric and acoustic properties are linear time-invariant. In practice, it is considered that air-flows vary acoustic properties in ducts and responses of the loudspeaker and the microphones will change slowly as passing time on.

It is considered to design the adaptive DSM system able to cope with the varying physical parameter. Fig.6 shows the adaptive DSM system and $H_e$ is updated using adaptive algorithms to minimize residual noises. The initial filter coefficients of $H_e$ may be estimated using LSA. In adaptive algorithms of the control system, two different algorithms, Least-Mean-Squares (LMS) algorithm and Recursive Least-Squares (RLS) algorithm [3] are available. The LMS algorithm is applied to the DSM system, because the RLS algorithm requires high computations.

The LMS update equation is following,

$$h_i(n) = h_i(n-1) + 2w(x(n-i)e(n)) \quad (3)$$

where $h_i(n)$ denotes a coefficients of the FIR filter, $x(n)$ a input signal of the filter, $e(n)$ a residual noise, $w$ a convergence factor, and $i$ a tap identification. Furthermore $M_2$ may be used to sense residual noises instead of $M_3$.

6. CONCLUSIONS

We described the principle, the configuration, the effect of the feedback reduction, the filter design of the DSM system, and the method to apply an adaptive algorithm to the DSM system. The system is constituted by four system components such as a FIR digital filter, a loudspeaker, and two microphones. Adaptive algorithms, therefore, are applied briefly to the DSM system because of its simple construction.

REFERENCES

A NEW APPROACH TO ACTIVE ATTENUATION IN DUCTS

L. J. Eriksson, M. C. Allie, and R. A. Greiner

Nelson Industries, Inc.
P. O. Box 428
Stoughton, WI 53589, U.S.A.

*Department of Electrical & Computer Engineering
University of Wisconsin-Madison
Madison, WI 53706, U.S.A.

INTRODUCTION

Active sound attenuation provides significant low frequency attenuation in a much smaller package size than is possible with passive silencers. A microphone is used to sense the undesired noise and re-introduce an inverted signal through a transducer to cancel the undesired noise. Several recent papers have provided reviews of the basic technology (1,2). More recently, efforts have focused on the use of adaptive digital signal processing to implement the controller used to generate the cancelling noise. System identification concepts have been used to formulate the basic problem, as shown in Fig. 1 (3). The physical plant is adaptively modelled such that the error signal or residual noise is minimized.

![Fig. 1 - System identification formulation of active attenuation problem](image1)

ACOUSTIC FEEDBACK

One of the most significant problems with active sound attenuation has been the presence of acoustic feedback from the secondary transducer to the input microphone, as shown in Fig. 2. In addition to the use of directional sound sources and microphones to minimize acoustic feedback, a variety of techniques have been used in an attempt to separately model the acoustic feedback path and to cancel the feedback signal in the same way that the primary noise is cancelled. However, these approaches have typically used a fixed feedback model that does not adapt to changes in the acoustic feedback path due to changes in such factors as temperature and flow rate.

![Fig. 2 - Acoustic feedback included as part of overall model](image2)

An alternative approach is to consider the feedback path as effectively part of the control model, as shown in Fig. 2. With this viewpoint it can be shown that the effect of the acoustic feedback is to introduce poles into the response of the overall model (3). The effect of these poles must be removed by a pole-zero or recursive infinite impulse response (IIR) adaptive filter model.

One practical method to use in implementing such an IIR adaptive filter is the recursive least mean square (RLMS) algorithm of Ref.1 (4), as shown in Fig. 3. It can be shown that this model is observable and well suited for use in an active attenuation system (3).

![Fig. 3 - RLMS algorithm](image3)

SPEAKER COMPENSATION

In order to use the RLMS algorithm it is necessary to compensate for the effects of the secondary transducer response characteristics. This may be done on an off-line basis with either a direct or inverse model of the sound source, as shown in Fig. 4 (3).

If the inverse modelling approach is chosen, the net effect of the speaker compensation process is to place a delay in the error path of the algorithm. This requires the addition of the same delay to the inputs of the error correlators for both sections of the RLMS algorithm.

If the direct modelling approach is chosen, the resulting model of the transducer response must be copied into the inputs of the error correlators for both sections of the RLMS algorithm.

![Fig. 4 - a) Direct speaker model b) Inverse speaker model](image4)
ERROR PATH COMPENSATION

Although it may be feasible to model the effects of the secondary sound source on an off-line basis, as described above, similar effects caused by a transfer function in the error path must be modelled on an on-line basis. This is due to the relatively rapid changes in the characteristics of the error plant that are possible with changes in system parameters such as temperature or airflow rate. Eriksson (3) has proposed the use of a third microphone to obtain an independent model of the error plant as shown in Fig. 5. The third microphone is located close to the secondary sound source. Although this nearfield location does not allow an error signal measurement that is sufficiently accurate for use in the primary RLMS model, it does allow a reasonable model to be obtained of the error plant. This approximate model of the error plant is adequate to ensure convergence of the primary RLMS model.

PERFORMANCE

A demonstration system shown in Fig. 6 has been implemented on a T.I. TMS32010 digital signal processing microprocessor. Results were obtained on an actual acoustic system with and without active attenuation. Figure 7 shows the relative sound pressure spectra with a pure tone input signal, while Fig. 8 contains similar results with a band-limited noise input signal. The system can adapt rapidly and produces up to about 30-80 dB of attenuation for narrowband signals and up to about 20-30 dB of attenuation for broadband signals over a significant frequency range.

CONCLUDING COMMENTS

This work has demonstrated an attractive approach to active attenuation that adaptively models both the direct and feedback plants as well as secondary transfer functions due to the error plant (5). Current research activities are concerned with improved methods for determination of the characteristics of the secondary sound source and error plant.

REFERENCES

5. Patents pending.
ELECTRIC SOUND CANCELLATION SYSTEM
FOR AIR-CONDITIONING DUCTS

M. Takahashi, T. Kuribayashi, K. Asami (1)
and H. Hamada, T. Enokiida, T. Miura (2)

(1) Hitachi Plant Engineering & Construction Co.,
Ltd. Research Laboratory
537 Kamihongo, Matsudo-shi, Chiba-ken, JAPAN
(2) Tokyo Denki University
2-2 Kandanishikicho, Chiyoda-ku, Tokyo, JAPAN

1. INTRODUCTION

The construction of air-conditioning duct systems in auditoriums, hospitals, and large buildings involves a precise prediction of airborne noise propagating through ductworks, plus compact but high-performance silencers to support it [1]. Increasing low-frequency attenuations with reactive silencers, however, is prone to many problems such as pressure drops and increased dimensions. Many researchers have recently focused on active silencers using digital signal processing techniques as alternative means without these problems [2], [3], [4], there exists, however, some problems in connection with the suppression of acoustic feedback for sensing microphones, adaptive control of silencer for the practical usefulness.

The purpose of this paper is to develop an electronic sound cancellation system for air-conditioning ducts using digital filters [5], [6]. We will propose an electronic sound cancellation system operating on dual sensing microphone (DSM) principles to suppress acoustic feedback, and report the results of its applications to broadband random noise and fan noise.

2. MODELLING OF ELECTRONIC SOUND CANCELLATION SYSTEM

In addition to noise propagation in ducts, correction of the characteristics of electric transducers, such as microphones and loudspeakers, is important to ensure maximum performance of the electronic sound cancellation system. In the general monopole system shown in Figure 1, correction of these characteristics is made possible by a model design as shown in Figure 2.

Where,
M1, M2: Microphones
S: Canceller (Loudspeaker)
P1, P2: Sound pressures
VA, VB, VC: Voltages at respective points
G: Transfer function defined by the sound pressures
Hr: Transfer function of VB between VA
Ht: Transfer function of VB between VC

If Vc = 0, the cancellation digital filter He is expressed by the following equation:

\[
He = \frac{Gd \times (HM2 / HM1)}{Ht - Gd \times (HM2 / HM1) \times He} \quad \ldots \ldots \quad (1)
\]

The transfer functions ofCd (HM2/HM1), Ht, and Hr necessary to determine He, are all identifiable at the measuring points VA, VB, and VC. The suppression of acoustic feedback from loudspeaker S to microphone M1 is essential to implementing a general monopole system.

3. DSM SYSTEM

The DSM system is a technique of synthesizing electrical signals for input to a cancellation digital filter, from signals obtained by dual sensing microphones. It is very simple in configuration. Figures 3 and 4 show the configuration and block diagram of the DSM system, respectively.

\[
\text{Fig. 1 General Monopole System Arrangement}
\]

\[
\text{Fig. 2 General Monopole System Block Diagram}
\]

\[
\text{Fig. 3 DSM System Configuration}
\]
Fig. 4 DSM System Block Diagram

On the basis of Figures 3 and 4, output VC of microphone M2 can be expressed by the following equation:

\[
VC = \frac{P1[H_1H_2 + G_A M_2(1 - H_2H_3)]}{1 - H_2(H_1 - H_3)} \quad \ldots (2)
\]

Transfer function \( H_2 \) of the cancellation digital filter is equal to the result of solving Eq. (1), it can be determined from Eq. (2) as VC = 0. Making \( H_1 \) and \( H_3 \) equal to each other cancels feedback from speaker S to microphone M1, with howling being suppressed. Though this condition is impractical to satisfy in precise terms, it can be practically accomplished by setting the microphones at equal distances from the cancellation loudspeaker in the duct [7].

4. APPLICATION TO VARIOUS SOUND SOURCES

4.1 Broadband Random Noise

The air-conditioning ducts used in the experiment were straight ducts measuring 350 mm x 350 mm and 13 m long. Figure 5 shows the results of applying the DSM system to the broadband random noise, generated by driving the noise-source loudspeaker with 20 Hz to 1 kHz white noise. The cancellation digital filter used was an FIR digital filter.

Fig. 5 Effects of Sound Cancellation for Broadband Random Noise

Attenuations were obtained over a broadband of 60 to 800 Hz, demonstrating essentially equal cancellation effects in the part of the duct downstream of microphone M2.

4.2 Fan Noise

Fan noise attenuations were measured by using a centrifugal fan (Q: 60 m³/min, P: 2450 Pa) as a noise source, and varying the in-duct flow velocity in the range 2.6-9.3 m/s, with the same cancellation digital filter transfer function as used in the above-described experiment, regardless of flow velocity. Figure 6 shows the attenuations measured at an in-duct flow velocity of 6.5 m/s.

Fig. 6 Effects of Sound Cancellation for Centrifugal Fan Noise, Flow Velocity 6.5 m/s.

While sound cancellation effects were noticeable over a broad band, a slight fall was observed in the attenuations with rapid increases in the flow velocity. We are also currently proceeding with research on the adaptive control of a sound cancellation system capable of following up changes in various transfer functions associated with flow velocity, internal pressure, temperature, and so on.

5. CONCLUSIONS

It was proposed that an electronic sound cancellation system for air-conditioning ducts operating on dual sensing microphone (DSM) principles to suppress acoustic feedback in an extremely simple setup. We wish to promote the ongoing research on the adaptive control of sound cancellation systems, to reach the stage of practical usefulness.

6. REFERENCES

ACTIVE ACOUSTIC IMPEDANCE

O.Kr.B. Pettersen, S. Sersdal

Acoustics Research Center, ELAB.
N-7034 Trondheim-NTH, Norway

INTRODUCTION

Both for noise control and room acoustic purposes, the use of active systems, including microphones, electronics and loudspeakers, are increasingly being used. Usually, an active system has the nature of a feedforward or feedback system and is, as such, very dependent on the acoustic environment where it is used. Each system has to be specifically adjusted for each application, putting severe restrictions to the installation procedure. Also, if not having an adaptive controller, the active system’s performance will depend very much on the long time stability of the acoustic environment. A more attractive system would, in many cases, be a locally reacting acoustic impedance. Once set to a specific value, it would retain this impedance independent of the acoustic conditions of the environment. So far, only the tight coupled monopole, described by Leventhall et al. //, has features of this kind.

By developing a more general and controllable active acoustic impedance, the problem of using an active electroacoustic system is reduced to specifying the acoustic impedance needed, whether it is for controlling the acoustics of a room or for noise reduction.

ACTIVE ACOUSTIC IMPEDANCE

One possible way of realising an active acoustic impedance is to mount a microphone close to the membrane of a loudspeaker and connect the two through a controllable transfer function $H_x$, as shown in Figure 1. This system has two main constraints; the transfer function $H_x$ has to be realisable for implementation using e.g. a FIR filter and the system has to be stable. These two factors limit the achievable range for the acoustic impedance, both in value and frequency. In our work as presented in this paper, we have tried to establish a method for finding this range.

The acoustic impedance $Z$ can be regarded as a transfer function between the particle velocity $u$ and the pressure $p$. When analysing the active system in Figure 1, the acoustic impedance, as seen into the loudspeaker, can be expressed as two cascade coupled transfer functions, Figure 2. $H_x$ is the acoustic impedance of the loudspeaker and would be measured with the active system turned off. $H$ in the feedback loop is proportional to $H_x$ and expresses the modification of the loudspeaker impedance obtained with the active system.

Figure 2. Signal flow diagram for the active acoustic impedance.

The most common loudspeaker type, the electrodynamic loudspeaker, has an acoustic impedance which is generally too high for many active noise reduction applications. The possible exception is at the mechanical resonance frequency when the damping is low, but in this case the frequency range is limited.

RESULTS

In the present study, we have aimed at an active system which lowers the acoustical impedance, makes it predominantly resistive and broadens the frequency range compared to the mechanical impedance of a passive loudspeaker.

Figure 3 shows some of the results. The upper diagram shows the impedance level $L_z + 20 \log \left| \frac{Z}{\mu} \right|$ for the loudspeaker (dashed line) and some of the possible impedance curves for active systems containing the same loudspeaker. The active systems comprise an ideal microphone and an ideal FIR filter having 25% coefficients. The loudspeaker has been given a $Q$ factor of 1 at the resonance frequency 175 Hz. The impedance level is normalised to the value $L_z$ of the loudspeaker at resonance.

The curves indicate that active systems may achieve low resistive acoustical impedance over a fairly broad frequency range.
Figure 3. Modulus and phase of some active acoustic impedances.
- Gashed line: Loudspeaker.
- Solid lines: Active systems. Nominal $L_0$: 0, -6, -12 and -18 dB.

REFERENCE
EXPERIMENTATIONS ET SIMULATION NUMÉRIQUE D'UN SYSTÈME D'ANTI-BRUIT.


Electricité de France, Direction des Études et Recherches, Département Acoustique
1, avenue du Général de Gaulle - 92141 Clamart Cédex FRANCE.

LE SYSTÈME ANTI-BRUIT EDF-CNRS :

Les bruits des systèmes de ventilation industriels ou d'échappement de moteurs sont riches en basses fréquences. Les silencieux classiques qui réfléchissent ou absorbent une partie de l'énergie sont efficaces en moyenne et hautes fréquences mais nécessitent des équipements de grandes dimensions pour atténuer convenablement les bruits en basses fréquences. Ils introduisent de ce fait des pertes de charge importantes dans les circuits.

L'anti-bruit ou absorption acoustique active constitue une réponse bien adaptée au traitement des nuisances en basses fréquences. Son principe, très simple, est décrit sur le schéma suivant :

![Diagramme du système anti-bruit](image1)

1 Système de détection amont-Upstream detection unit
2 Filtre permettant de minimiser la fonction de transfert entre 1 et 4 Inversion de signal - Témoing filter to minimize filter signal inversion
3 Source de contrôle-bruit - Anti-noise source
4 Système de contrôle aval - Downstream checking unit

Figure 1 : Schéma de principe du système EDF-CNRS. - The principle of the EDF-CNRS system

Le système de commande ou "filtre" constitue l'élément essentiel du procédé et c'est autour de lui que se sont principalement concentrées les recherches engagées depuis quelques années. Une collaboration avec le CNRS-LMA a abouti à la mise au point d'un système possédant une efficacité importante sur une large bande de fréquence. Ce système utilise un filtre numérique non récursif programmable du type "convoluteur". Il est muni d'un échantillonneur d'entrée, permettant un fonctionnement en temps réel et une adaptation automatique aux variations de conditions de fonctionnement.

EXPERIMENTATIONS :

Dans un premier temps un système de laboratoire a été réalisé. Un test sur un conduit d'aérofrigeration d'un poste de transformation dans un immeuble à Paris a démontré la possibilité d'application à une installation industrielle (cf. figures 2, 3 et 4).

(*) Centre National de la Recherche Scientifique. Laboratoire de Mécanique et d'Acoustique de Marseille.

(**) Poste de Transformation EDF de Cardinal-Lemoine

Les résultats ont été excellents puisqu’au voisinage du poste on a observé une attenuation de 20 dB autour de 100 Hz (fréquence du ventilateur) et qu'à l'intérieur du poste l'atténuation était en moyenne de 15 dB sur la bande de fréquence allant de 80 à 250 Hz. Le conduit traité a un diamètre de 1 m et la vitesse d'écoulement est de 5 m/s.

Les récents travaux ont eu pour but de mettre au point un prototype miniaturisé "industriel". Les études ont abouti à la réalisation d'un système intégré piloté par un micro-ordinateur. Il détermine les paramètres du filtre numérique et les actualise périodiquement. Le remplacement de ce calculateur par deux cartes électroniques intégrées dans le prototype actuel est en cours. D'autre part l'élément essentiel du filtre est constitué d'un processeur spécialisé dans le traitement du signal du type TMS 320.10.

Ce prototype miniaturisé (cf. figure 5) a également été testé en conditions de fonctionnement industrielles. Une installation d'essais a été réalisée sur un second poste en immeuble à Paris. Les résultats obtenus sont concluants puisque le pic de bruit du ventilateur à la fréquence de 164 Hz est effacé : seul subsiste le bruit de fond de l'installation (cf. figure 6).

(*) Poste de Transformation EDF de Faidherbe.
Ces résultats prouvent la faisabilité d'un système d'anti-bruit "industriel". Le développement technologique est pratiquement achevé : quelques études tendant à améliorer les performances ou la fiabilité du système sont en cours. Ce procédé peut maintenant être industrialisé.

. SIMULATION NUMERIQUE DE SYSTEMES D'ANTI-BRUIT.

En appui des développements technologiques, des outils informatiques ont été mis au point afin de mieux comprendre le phénomène d'absorption acoustique active et d'améliorer les performances des systèmes anti-bruit. C'est ainsi qu'un code temporel résolvant l'équation des ondes dans un conduit pouvant posséder des sources en paroi a été développé. Ce programme de calcul utilise une méthode aux différences finies explicite pour résoudre les équations suivantes:

\[
\frac{\varphi(t)}{C^2 \partial t^2} - \frac{\partial^2 \varphi}{\partial x^2} = 0 \quad \text{dans le conduit}
\]

\[
\frac{\partial \varphi}{\partial n} = \psi(s) \quad \text{au niveau d'une source}
\]

\[
\frac{\partial \varphi}{\partial n} + \frac{A_k}{C} \frac{\partial \varphi}{\partial t} = k_0 A_{A_k} \varphi \quad \text{sur toute paroi}
\]

\[
\text{d'admittance efficace } A_k = A_k + \frac{1}{A_k}
\]

\[
\varphi(t=0) = 0 \quad \text{à l'inst ant initial}
\]

\( \varphi \) désigne le potentiel de vitesse. Il est relié à la pression acoustique \( p \) et la vitesse vibratoire \( \mathbf{v} \) par les relations:

\[
\begin{align*}
\mathbf{v} &= \text{grad} \varphi \quad (\text{fluide isotrope}) \\
\rho \frac{\partial \varphi}{\partial t} &= \varphi \quad (\varphi = \text{masse volumique du fluide})
\end{align*}
\]

La connaissance de \( \varphi \) suffit donc pour déterminer le champ acoustique instantané. Un moyen temporel permet d'obtenir les valeurs efficaces en pression et vitesse ainsi que l'intensité acoustique.

Ce code de calcul a été utilisé pour simuler deux systèmes d'anti-bruit. Le premier utilise une seule contre-source alors que le second utilise un couple de contre-sources directif. Les simulations effectuées sont illustrées par les figures suivantes:

Il apparaît clairement que les deux systèmes sont très bien simulés:
- pour le système mono-source on retrouve l'onde stationnaire entre le haut-parleur primaire et le haut-parleur secondaire due à l'interaction entre l'onde qui "descend" le conduit et l'onde issue de la source secondaire qui "remonte" le conduit.
- pour le système bi-source ce phénomène n'existe plus du fait de la directivité du doublet de sources : aucune onde n'est générée vers l'amont par les hauts-parleurs secondaires et la source primaire rayonne donc normalement jusqu'au doublet.

On peut noter la complémentarité des résultats intensimétriques et des résultats en pressions. Ceux-ci ne font pas apparaître de grosses différences de fonctionnement entre les deux systèmes. Par contre, les cartes intensimétriques renseignent beaucoup plus sur ce qui se passe dans le conduit et au niveau des sources:
- dans le système mono-source l'intensité est pratiquement nulle dans tout le conduit. L'anti-bruit modifie le fonctionnement du haut-parleur primaire : celui-ci ne rayonne pas plus alors qu'il a toujours la même vitesse vibratoire. Il ne génère pratiquement pas d'énergie acoustique dans le conduit.
- dans le système à deux contre-sources, la source primaire rayonne normalement. L'énergie acoustique générée est littéralement absorbée par le doublet de sources d'où la justification du terme d'absorbeur actif donné à un tel système.

Ces résultats démontrent l'intérêt d'une simulation numérique des phénomènes intensimétriques car ils mettent en évidence des flux d'énergie difficilement mesurables susceptibles de modifier le dimensionnement des sources utilisées dans les systèmes anti-bruit.
ACTIVE NOISE ATTENUATIONS IN THREE-DIMENSIONAL SPACE

Sha Jia-zheng and Tian Jing
Institute of Acoustics, Nanjing University, Nanjing, China

INTRODUCTION

It is well known that an active noise attenuator includes the main parts: (1) primary noise signal pick-up; (2) an electronic system; (3) secondary sound radiation sources. Since the second part (analogous or digital ones) of Active Noise Attenuations (ANA) has been studied very well in many of the published papers, more attention to the experimental realization of secondary sound radiation sources and an available way of primary noise signal pick-up. At present stage, they are the only two outstanding problems that have to be solved in experiments, especially for the ANA in three-dimensional space. Moreover, the exploration of acoustical absorption mechanism of ANA is also an interesting and practical problem. Here, we try to give an explanation on the basis of laboratory studies.

SECONDARY SOURCES

When a primary noise source can be supposed as a point source which radiates a pure-tone of low frequency, a theoretical expression of sound radiation characteristics of a discrete secondary source of ANA in three-dimensional space can be given by the Kirchhoff's formula and sound field diffraction theory:

\[ P(x,y,z) = \frac{A'}{r} \left( \text{e}^{ikr} + jkr \cos \theta \right) \exp[-jk(r+r')] \]

\[(kr > 1)\]

where the parameter \( r', r \) and \( \theta \) are shown in Fig.1. \( A' \) is a constant depending on the relative solid angle controlled by the secondary source and the amplitude of primary sound pressure of where the secondary source is located.

The physical meaning of the formula is that, under the far-field condition of \( kr > 1 \), a discrete secondary source has the sound radiation characteristics of a given monopole, and both the primary phase and amplitude of the monopole are modulated by frequency. In the formula, every parameter can be compensated through the electronic system composed of amplifiers, time-delay and phase-shifter, etc., except the directivity parameter \( \cos \theta \). So, only this term must be considered from acoustical designs.

A source composed of a monopole and a phase shifted dipole is given in Fig.2. A simple theoretical analysis has shown that its characteristics of sound radiation are well content with that demanded by a discrete secondary source. Practical performances are perfect, too. Fig.3 gives an experimental result of a pure-tone noise absorption when such a triple secondary source is used, from which we can see that no reinforcement of primary noise field appears in the whole space.

* Projects Supported by the Science Fund of the Chinese Academy of Sciences.

EXPERIMENTAL APPROACH

Experiments were carried out in an anechoic chamber and a reverberation room. The volume of the room was 7.37m$^3$. Fig.3 shows that a single secondary source has got a noise attenuation in all space already, but the reduction is quite little out of a limited space region. So in the experiments, four of the designed triple secondary sources were arranged around the primary noise source imitation.

In order to decrease the effect of secondary sound signal feedback, instead of using a primary noise signal picked up by a microphone as the input of an ANA system, a vibration signal got by a small accelerometer that sticked to the radiation surface of the primary source imitation was used. Experiments had proven the feasibility of the replacement when the operating frequencies were low enough to ensure the radiation surface not conducting a damped vibration.

To see if the primary noise energy is absorbed or just reflected by the secondary sources, we measured not only the whole spatial sound fields before and after the attenuation, but also the variation of the motional impedance of a secondary source through the measuring of its vibration velocity. The experimental set-up is shown in Fig.4.

EXPERIMENTAL RESULTS AND DISCUSSIONS

Experiments had shown that in the frequency range of 50-400 Hz, noise reductions were achieved in all the space out of the closed surface formed by secondary sources both in the anechoic chamber and in the reverberation room. In the chamber, the optimum attenuation indices were 34dB for pure-tones, 22dB for 1/3 oct. band-noises and 16dB for 1 oct. band-noises. In the room, these values were 28dB, 17dB and 12dB respectively. Typical polar sound level distributions in a horizontal plane of the chamber before and after the reduction are shown in Fig.5. A broadband pink noise reduction curve is shown in Fig.6, which corresponds to a total sound level attenuation of 10 dB and an effective frequency range of 65-350 Hz.

The sound field measurements told us that the primary noise energy was not reflected by secondary sources, because the noise level attenuation was achieved everywhere in the whole field. On the other hand, when the primary source and secondary sources operated simultaneously and in a steady noise attenuated working state, the vibration velocity of the secondary sources also got a higher level compared with that as the secondary sources worked alone (Fig.4). This may be interpreted by an increment of motional impedance of secondary sources to which was led by a negative value of their radiation resistance. So it seemed likely that the primary noise energy was absorbed by the secondary sources. This conclusion is well content with the theoretical explanation presented by N. Jessel.

CONCLUSIONS

The main points of the experimental study can be summarized as follows:

1. The design of proper secondary sources is a key problem in the study of real spatial active noise attenuators. Triple secondary sources presented in this paper have given perfect practical performances.
(2) The energy transfer in the active noise attenuations in three-dimensional space is still a problem worthy to be studied further. It all depends on the radiation characteristics of the used secondary sources that the primary noise is absorbed or just reflected. In our experiments, the designed dipole source was a sort of energy absorbent type, so the sound field attenuations could be obtained even in a reverberation room. For a sound energy reflective type of secondary sources, it is impossible.

(3) The experiments were conducted not only in an anechoic chamber, but also in a reverberation room. So the obtained results have some general significance in practical applications.

REFERENCES


Fig.1 Diagram of the parameters of the formula

Fig.2 A realization of discrete secondary sources

Fig.3 A practical performance of the designed secondary source

Fig.4 Experimental set-up for measuring the vibration velocity
1—Oscillator, 2—Phase shifter, 3,4—Power amplifier, 5—Charge amplifier, 6—Accelerometer.

Fig.5 Polar sound level recordings in the anechoic chamber

Fig.6 Attenuation of a broadband noise

PSL: Primary Sound Level  RSL: Reduced Sound Level
ACOUSTIC PULSATION OF PIPING SYSTEMS BY THE TRANSFER MATRIX AND FINITE ELEMENT METHODS

C.W.S. To

Department of Mechanical Engineering, The University of Calgary, Calgary, Alberta, Canada, T2N 1N4

1. INTRODUCTION

Designers of reciprocating compressor installations and other piping systems are often confronted with the problem of acoustic pulsation generated by the compressor(s). When the size and complexities of the installation are large and the operating pressure is high the problem can be extremely formidable. Consequently, a digital computer analysis of the acoustic system can be used to examine possible major vibration problems at the design stage or to reduce such problems of existing systems to an acceptable level. Currently, digital computer programs available for the acoustic analysis of piping and similar systems are based on the transfer matrix approaches (TMA) [1-6], and the finite element methods (FEM) [7-9]. While the finite elements in references [7,8] can deal with three-dimensional acoustic problems and therefore more versatile, it is generally agreed that the one-dimensional acoustic model based on the TMA and the FEM is adequate for the analysis as, frequently, the transverse dimensions of the piping systems are small compared with the wavelength. Moreover, the cost of acoustic analysis for a typical piping system based on the finite element representation of the three-dimensional acoustic model can easily exceed an order of magnitude higher than that of an one-dimensional finite element model. Naturally, the question to be asked at this stage is, "Which one-dimensional digital computer program should an analyst choose for the analysis of acoustic pulsation in piping systems?"

The main theme of this paper is therefore, concerned with an attempt to answer the aforementioned question by way of analysing a typical piping system using the program, PASAPS I [1,2], and another program based on the finite element formulation of reference [9]. The program, PASAPS I, has been chosen in this study for its comprehensiveness and user friendliness while the formulation in reference [9] has been selected for its simplicity and availability.

The intermediate following two sections shall briefly introduce the TMA and the one-dimensional FEM. Section 4 includes some results of the analysis. Concluding remarks are presented in Section 5.

2. TRANSFER MATRIX APPROACH (TMA)

The detailed formulation of the TMA adopted for the digital computer program, PASAPS I, has recently been proposed by the author [1]. Briefly, any large and complicated piping system is represented by a finite number of elementary acoustic components. Every acoustic component is described by the one-dimensional distributed-parameter model and possesses two stations. In the frequency domain, the acoustic pressure and volume velocity at one station are related to those at the other station by a two-by-two parameter matrix. Any part of the piping system can be modelled by an appropriate combination of parameter matrices of uniform pipes of various cross-section areas.

mean flow effect has been included while the temperature effect can also be represented by considering various sound speeds and densities of the moving medium for an appropriate number of elementary components.

In this TMA the acoustic resonant frequencies of the piping system are defined as those corresponding to absolute maxima of the transfer impedance

\[ Z_{ji} = \frac{P_j}{V_i} = \frac{(Z_1a_{11} - a_{12})/|a|}{(Z_2a_{21} + a_{22})}, j = 2,3,\ldots \]

where \( Z_1 \) = \((Z_1a_{11} + a_{12})/(Z_1a_{21} + a_{22}) \), \( a_1 \) is the acoustic volume at station \( j \).

\( P_j \) is the acoustic pressure (in the frequency domain) at station \( j \); \( Z_j \) is the acoustic impedance at station \( j \); \( a_j \) is the four-pole parameter matrix of the acoustic component or element associated with station \( j \).

Applying Equation (1) and giving the acoustic volume velocity of the source the acoustic pressure at every station can be computed.

3. FINITE ELEMENT METHOD (FEM)

In this method the acoustic pressure or velocity over an element is assumed to vary proportionally with element length [9]. The acoustic equations for a piping system can then be expressed as

\[ [N][\ddot{p}] + [C][\dot{p}] + [K][p] = [f], \]

where \([p] \) is a vector of nodal pressures while \([f] \) is the volumetric acceleration injected at every node; the matrices \([M], [C], \) and \([K] \) are respectively formed from the individual element mass, damping, and stiffness matrices. They are included in the appendix.

Taking the Fourier-transformation of Equation (2) results

\[ ([K] + i\omega[C] - \omega^2[M])\dot{p} = i\omega[V], \]

where \( i = (-1)^{1/2} \). \([p] \) and \([V] \) are the Fourier-transformation of \([p] \) and injected volume velocity vector respectively.

Dividing both sides of Equation (3) by the volume velocity at station \( 1 \), \( V_1 \), and rearranging one has

\[ \begin{bmatrix} Z_{11} \\ Z_{12} \end{bmatrix} = \text{iu}([K] + i\omega[C] - \omega^2[M])^{-1} \begin{bmatrix} 1 \\ V_1 \end{bmatrix}, \]

in which the elements of the vector on the left-hand side of Equation (4) are the transfer impedances and \( V_{11} = V_1/V_1 \), etc.

The frequencies corresponding to the absolute maxima of Equation (4) are the acoustic resonant frequencies of the system. Note that when the term inside the parentheses on the right-hand side of Equation (4) is set equal to zero it gives the resonant frequencies of the system. This reflects that the definition of resonant frequencies of the system based on Equation (1) is correct.

Using Equation (2) - (4), the acoustic pressures at the nodes, acoustic resonant frequencies and acoustic impedances at various station can be obtained.
4. RESULTS

The typical piping system investigated is the one considered in reference [2] and its diagrammatical representation is included in Figure 1. For brevity, selected computed resonant frequencies are presented in Table 1. It should be noted that in the TMA a total of 39 acoustic elements (components) were employed to represent the system in Figure 1 while in the FEM 280 individual finite elements (281 degrees of freedom) were required to approximate the same system. Moreover, the time for data input and analysis using the FEM is about 5 times longer than that with the TMA.

5. CONCLUDING REMARKS

The resonant frequencies obtained with the FEM are upper bounds to those employing the TMA. The FEM required much more computer storage than the TMA. Furthermore, the time for the FEM pre-processing and analysis is much longer than that necessary for the TMA. Thus, the TMA is much more economical to employ compared with the FEM.

REFERENCES


APPENDIX

The acoustic mass, damping, and stiffness matrices are given below.

The element mass matrix is

\[ [m] = \frac{A\ell}{60(\rho c^2 - \mu^2)} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \tag{A1} \]

where \( \ell \) - element length,
\( A \) - cross-sectional area of the element,
\( c \) - sound speed of the medium in the pipe,
\( U \) - steady mean flow speed of the medium, and
\( \rho \) - density of the medium.

The element stiffness matrix is

\[ [k] = \frac{A}{\rho \eta} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{A2} \]

The element damping matrix is

\[ [c] = [c_f] + [c_c], \tag{A3} \]

where

\[ [c_f] = \frac{3A}{\rho U c(\rho c^2 - \mu^2)} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \]

\[ [c_c] = -\frac{A}{\rho (\rho c^2 - \mu^2)} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}; \]

\( (\Delta p/L) \) is the static pressure drop per unit length of the pipe. This pressure loss term can be analytically estimated using a Darcy-Weisbach friction model.

To model an orifice and anechoic end condition two different damping matrices are required. They can be found in reference [9] and shall not be included here for brevity.

---

Figure 1. First stage suction of a gas compressor station. FSSPB, first stage suction pulsation bottle; FSSS, first stage suction scrubber; DMT, Daniel meter tube; BCV, body control valve.

<table>
<thead>
<tr>
<th>Mode number</th>
<th>TMA</th>
<th>FEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4</td>
<td>4.1</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>12.5</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>16.1</td>
</tr>
<tr>
<td>4</td>
<td>24</td>
<td>26.0</td>
</tr>
<tr>
<td>5</td>
<td>27</td>
<td>28.3</td>
</tr>
<tr>
<td>6</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>36</td>
<td>38.7</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Resonant frequencies of the piping system in Figure 1.
DISPERSE NONLINEAR WAVE INTERACTIONS IN A RECTANGULAR DUCT

J. A. TenCate and M. F. Hamilton

Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029, and Department of Mechanical Engineering, The University of Texas at Austin, Austin, Texas 78712-1063 USA

Introduction

Experiments are described in which finite amplitude sound in the plane wave mode (0,0) and first higher order mode (1,0) is generated in a hard wall rectangular duct. Two types of nonlinear wave problems are considered. The first involves the weak nonlinear interaction of a wave of (angular) frequency ω1 in the (1,0) mode with a wave of frequency ω2 in the (0,0) mode. Because the phase speeds of the waves in the (0,0) and (1,0) modes are very different, the nonlinear interaction is strongly affected by dispersion. Results are presented for the levels of the nonlinearly generated waves at the sum (ω1 + ω2) and difference (ω1 - ω2) frequencies. In the second problem we investigate the finite amplitude distortion of a pure tone generated in the (1,0) mode. For a certain range of frequencies, the distortion of sound in the (1,0) mode is relatively unaffected by dispersion. The propagation of frequencies in this range may be described by modifying slightly the theory for finite amplitude plane waves. Data obtained from both the bifrequency and single frequency experiments are compared with analytical results. Additional experimental results and details of the analytical work will appear in forthcoming papers by the authors.

Experiments

The rectangular aluminum duct used for the experiments has a length of 8 m, a wall thickness of 0.3 cm, and an inside cross section (a x b) of 7 cm x 3.8 cm. A diagram of the apparatus is shown in Fig. 1. Since only progressive waves are desired, one end of the duct contains an anechoic termination, a 2 m long tapered wedge of batted Kevlar.

An acoustic source configuration that generates sound in both the (0,0) and (1,0) modes simultaneously is essential to the bifrequency experiment. The (0,0) mode is generated with a JBL compression driver that is mounted directly to the end of the duct. Excitation of the (1,0) mode is achieved with two University Sound ID-65 drivers mounted on the top and bottom walls (see Fig. 1) and driven 180° out of phase. Sound from each driver enters the duct through a narrow (0.3 cm x 3.5 cm) slot machined in the wall. Each slot is parallel to the y axis and located approximately 14 cm from the JBL diaphragm. For the second experiment the JBL driver is removed and an aluminum plug is inserted in its place. For air at room temperature, the cutoff frequency of the (1,0) mode is about 2700 Hz.

Measurements of the sound pressure are made along the length of the duct. Microphone port holes are located along the top wall at distances out to 5.5 m. The distance between ports varies but is never more than 30 cm. Each hole accepts a B&K 4136 1/4 in. microphone (no protective grid) and Teflon adapter designed so that the adapter and microphone diaphragm are flush with the inside wall. The ports are stopped up when not in use. Two microphones are used in the experiments. One is placed in the microphone port nearest the sources (7 cm downstream from the ID-65 drivers) and is used to monitor source levels. The other is used to make measurements at the remaining ports.

Theory

For the first experiment a quasilinear solution is derived for the sum and difference frequency pressure p. It is assumed that the nonlinear interaction between the two waves is weak. The analysis is similar to that performed by Novikov et al. 2 for the nonlinear interaction of two plane waves in a lossless fluid. However, we have modified their analysis for application to waves in a hard wall rectangular duct and have introduced attenuation ad hoc. The primary wave field may be described by

\[ p = p_{01} \cos(\omega_1 x) \sin(\omega_1 t - k_1 z) \cos(\theta_1) \exp(-\alpha_1 z) + p_{02} \sin(\omega_2 t - k_2 z) \exp(-\alpha_2 z), \]

where \( p_{01} \) is the amplitude along the wall of the primary wave (1,0), \( k_1 = \omega_1/c_0 \) the wavenumber, \( \alpha_1 \) the attenuation coefficient, and \( c_0 \) the sound speed. The angle \( \theta_1 = \arcsin(\omega_1 k_1) \) and the coordinates \( x \) and \( z \) are defined in Fig. 1. The quasilinear solution for the complex sum and difference frequency amplitude \( P_k \) is

\[ P_k = \pm (2 \omega_2^2 P_{01} P_{02} / \rho_0^2 c_0^2) \cos(\pi x/a) \exp(-\Delta_0 z/2), \]

where \( \omega_d = \omega_1 + \omega_2, \rho_0 = \text{Im}(P_{01} \exp(i \omega_0 t)), \) and the boundary condition \( P_k (x = 0) = 0 \) has been applied. The factors

\[ \Delta_0 = (k_1 c_0 \theta_1 + k_2 c_0 \theta_2 - j \omega_0) \]

include the angle \( \theta_2 = \arcsin(\omega_2 k_2) \), the wavenumber \( k_2 = \omega_2/c_0 \), and the attenuation coefficient \( \alpha_2 \) associated with the nonlinearly generated wave at frequency \( \omega_2 \). The factor \( \beta_1(\theta_1) = (B/2A) \cos \theta_1 + 4(\omega_1 \omega_2 / \omega^2) \sin^2(\theta_1/2) \) is a modified coefficient of nonlinearity, with B/A a ratio of coefficients in the isentropic equation of state relating the pressure and density. From Eq. (2) it is seen that both the sum and difference frequency waves are generated in the (1,0) mode.

The second experiment involves the distortion of a finite amplitude pure tone generated in the (1,0) mode. In this case, three frequency ranges may be identified. One, near very cutoff, is characterized by standing waves formed across the duct. Very little energy is propagated down the duct. A second, at frequencies far above cutoff, is characterized by moderate dispersion. All components of the nonlinearly distorted wave propagate down the duct with approximately equal phase speeds. The third is in a mid-frequency range where dispersion is sufficiently strong that only waves propagating with identical phase speeds are strongly coupled. Waves in a rectangular duct propagate with identical phase speeds when the ratio \( m/a \) is maintained, where \( m \) is the index denoting the order of the (m,0) mode. Thus, if a source launches a wave of frequency \( \omega_1 \) in the (1,0) mode, the harmonic of frequency \( \omega_1 \) propagates at the same phase speed in the (n,0) mode. If we consider only the interaction between waves with identical phase speeds, a solution for the pressure may be written in the form

\[ p(x,z,t) = \sum_{n=-\infty}^{\infty} p_n (z) \cos(n \pi x/a) \exp\left[i n(\omega_1 t - k_1 z \cos \theta_1)\right], \]
where \( |p_n| \) is half the amplitude of the pressure that would be measured for the \( n \)th harmonic at the wall of the duct. Solutions for the complex quantities \( p_n \) are found from the coupled set of equations \((-\infty < n < \infty)\)

\[
\frac{dp_n}{dz} + \alpha_n p_n = (j\omega_1/4 \rho c \cos \theta_1 X_1 + 8/2A) \sum_{i=-\infty}^{\infty} p_i p_1
\]

Equation (4) is a modification of Korpel's solution for plane waves, where \( \alpha_1 \) is the attenuation coefficient of the \( n \)th harmonic, and \( \theta_1 = \arcsin (\pi/k_1) \) is again the angle associated with the wave in the \((1,0)\) mode. From the right hand side of Eq. (4) it is seen that the accumulation of nonlinear effects in the \( z \) direction is proportional to \( 1/\cos \theta_1 \). For the case in which the source pressure is \( A \cos (\pi x/2 \sin \theta_1) \), we then have at \( z = 0 \): \( p_{-1} = jA/2 \), \( p_1 = -jA/2 \), and \( p_{n} = 0 \) for \( |n| \neq 1 \).

**Results**

In the first experiment, the two frequencies are \( f_1 = 3200 \) Hz \((1,0)\) mode) and \( f_2 = 16.3 \) Hz \((0,0)\) mode). In this case the angle \( \theta_1 \) between the propagation directions of the two waves is \( 50^\circ \). The sound pressure levels (in dB re 20 \( \mu \)Pa) of the two primary waves at \( z = 0 \) (the position of the monitor microphone) are, respectively, 120.7 dB and 129.8 dB. The source levels are sufficiently low that no significant finite amplitude losses are incurred by the primary waves. In Fig. 2 are plotted the sum and difference frequency data (sound pressure levels). All levels are measured along the inside wall as explained above. Also shown are predictions obtained from Eq. (2). For these curves values for the attenuation coefficients are found using the theory of Beatty. The only adjustment made in plotting the theoretical curves in Fig. 2 was the choice of origin of the nonlinear interaction. Since the physical sources of the primary waves were separated by about 0.14 m, the actual origin of the nonlinear interaction was uncertain. When plotting the curves shown in Fig. 2, we assumed the origin to be at \(-0.1\) m (the slots for the \( \theta = 5^\circ \) drivers are located at \( z = -0.07\) m and the JBL diaphragm is located at \( z = -0.21\) m). Regardless, the periods of the spatial oscillations are in excellent agreement with theory. The slight discrepancies in amplitude probably result from an inability to measure the precise boundary condition for \( p_1 \) at the theoretical origin of the interaction. Recall that in deriving Eq. (2) it is assumed that \( p_{-1}(z = 0) = 0 \), whereas in fact it may be nonzero because of local nonlinear effects.

In the second experiment, a pure tone of frequency \( f_1 = 2900 \) Hz and level of 140 dB was generated in the \((1,0)\) mode. The measured pressures of the fundamental through fifth harmonic along the length of the duct are shown in Fig. 3. Data obtained at \( z = 0 \) provided boundary conditions for Eq. (4), while attenuation coefficients were again calculated using Ref. 5. Theory is plotted together with the data, and the agreement for the first through the fourth harmonic is excellent. As the fundamental frequency departs from the vicinity of 2900 Hz, agreement between theory and experiment deteriorates.

**Acknowledgments**

The authors would like to thank D. T. Blackstock for helpful discussions as well as providing the initial motivation for this work. The support of the U.S. Office of Naval Research is also gratefully acknowledged.

**References**

A NUMERICAL METHOD FOR SOUND PROPAGATION IN THREE-DIMENSIONAL RECTANGULAR DUCTS

Zhu Zhichan and Wang Zhiqiu

Department of Engineering Mechanics
Tsinghua University
Beijing, China

INTRODUCTION

Sound propagation in a duct is a classical and practical acoustic problem. A lot of finite difference solutions about it have been presented since 1970s[1]. They come from two kinds of methods. One is the steady state or the transient moment method to solve the Helmholtz equation for the acoustic pressure. The other is the transient finite difference method to solve the wave equation for the acoustic pressure. In both methods, the derivatives are represented by the usual central difference in space and time. However, there exist some difficulties at the present. The former method needs large matrix storage while the latter could not converge for some circumstances, especially for cut-off modes of the sound sources. Furthermore, the example of the numerical result for sound propagation in 3-D ducts has not been seen by us yet.

The MacCormack method which has two step schemes, i.e. predictor and corrector, was presented for solving the equations of compressible flow in 1969[2] and the Navier-Stokes equations in 1981[3], but it has not yet been applied to sound propagation problems for the difficulties of approach to the acoustic boundary conditions.

A number of efforts have been made to use the MacCormack explicit method for the linearized governing equations for the acoustic pressure and velocities and to approach to the acoustic boundary conditions in terms of normal characteristic analysis perfectly for the problem of sound propagation in ducts. Unfortunately, no convergent solution could be got in this way. However, dealing with the acoustic boundary conditions according to governing equations, characteristic analysis, physical concepts and numerical technique rather than characteristic analysis only and trying them through a lot of numerical tests, we have got the convergent solutions successfully.

In this paper, the MacCormack time-split explicit method which divides a 3-D problem into three 1-D ones, i.e. eqn.(1) can be rewritten equivalently as follows:

\[ \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0, \quad \frac{\partial p}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad \frac{\partial u}{\partial t} + \frac{\partial v}{\partial y} = 0, \quad \frac{\partial v}{\partial t} + \frac{\partial w}{\partial z} = 0. \] (4)

The MacCormack explicit finite difference schemes can be given as:

\[ \begin{align*}
    \bar{u}_{ijk}^{n+1} &= \bar{u}_{ijk}^{n} - \frac{\partial}{\partial x} \left( \frac{\partial u}{\partial x} \bar{u}_{ijk}^{n} - \bar{u}_{ijk}^{n} \right) \quad (6)
\end{align*} \]

\[ \begin{align*}
    \bar{v}_{ijk}^{n+1} &= \bar{v}_{ijk}^{n} - \frac{\partial}{\partial y} \left( \frac{\partial v}{\partial y} \bar{v}_{ijk}^{n} - \bar{v}_{ijk}^{n} \right) \quad (7)
\end{align*} \]

\[ \begin{align*}
    \bar{w}_{ijk}^{n+1} &= \bar{w}_{ijk}^{n} - \frac{\partial}{\partial z} \left( \frac{\partial w}{\partial z} \bar{w}_{ijk}^{n} - \bar{w}_{ijk}^{n} \right) \quad (8)
\end{align*} \]

\[ \begin{align*}
    \bar{p}_{ijk}^{n+1} &= \bar{p}_{ijk}^{n} - \frac{\partial}{\partial x} \left( \frac{\partial p}{\partial x} \bar{p}_{ijk}^{n} - \bar{p}_{ijk}^{n} \right) \quad (9)
\end{align*} \]

\[ \begin{align*}
    \bar{u}_{ijk}^{n+1} &= \bar{u}_{ijk}^{n} + \frac{\partial}{\partial x} \left( \frac{\partial u}{\partial x} \bar{u}_{ijk}^{n} - \bar{u}_{ijk}^{n} \right) \quad (10)
\end{align*} \]

\[ \begin{align*}
    \bar{v}_{ijk}^{n+1} &= \bar{v}_{ijk}^{n} + \frac{\partial}{\partial y} \left( \frac{\partial v}{\partial y} \bar{v}_{ijk}^{n} - \bar{v}_{ijk}^{n} \right) \quad (11)
\end{align*} \]

\[ \begin{align*}
    \bar{w}_{ijk}^{n+1} &= \bar{w}_{ijk}^{n} + \frac{\partial}{\partial z} \left( \frac{\partial w}{\partial z} \bar{w}_{ijk}^{n} - \bar{w}_{ijk}^{n} \right) \quad (12)
\end{align*} \]

Equations (5) are the finite difference schemes for interior points, which are second-order accurate in space and time. The finite difference schemes for boundary points are given as:

at entrance (i=1):

\[ p_{i-1,j,k} = \exp(2kix)\bar{p}_{i,j,k}^{n-1}, \]

\[ u_{i,j,k} = (\bar{u}_{i,j,k}^{n-1} - \bar{u}_{i-1,j,k}^{n-1})/\Delta x, \]

\[ v_{i,j,k} = (\bar{v}_{i-1,j,k}^{n-1} - \bar{v}_{i,j,k}^{n-1})/\Delta y, \]

\[ w_{i,j,k} = (\bar{w}_{i,j,k}^{n-1} - \bar{w}_{i,j-1,k}^{n-1})/\Delta z, \]

\[ p_{i,j,k} = \bar{p}_{i,j,k}^{n} - \frac{\partial}{\partial x} \left( \frac{\partial p}{\partial x} \bar{p}_{i,j,k}^{n} - \bar{p}_{i,j,k}^{n} \right) \quad (13) \]

\[ \frac{\partial p}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0, \quad \frac{\partial p}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad \frac{\partial v}{\partial t} + \frac{\partial u}{\partial y} = 0, \quad \frac{\partial v}{\partial t} + \frac{\partial u}{\partial y} = 0, \quad \frac{\partial w}{\partial t} + \frac{\partial u}{\partial z} = 0. \]

NUMERICAL METHOD

The MacCormack time-split method divides a 3-D problem into three 1-D ones, i.e. eqn.(1) can be rewritten equivalently as follows:

\[ \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = 0, \quad \frac{\partial p}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad \frac{\partial u}{\partial t} + \frac{\partial v}{\partial y} = 0, \quad \frac{\partial v}{\partial t} + \frac{\partial u}{\partial y} = 0, \quad \frac{\partial w}{\partial t} + \frac{\partial u}{\partial z} = 0. \]

\[ (2) \]

where \( M \) is the Mach number of the uniform flow velocity in a duct and \( f \) is the dimensionless frequency of the sound source.

Boundary Conditions

Physical boundary conditions are obtained on bases of characteristic analysis. For simplicity, here are the boundary conditions for sound propagation in a duct for \( \omega \) and \( \phi \):

at entrance, \( x=0, p=\exp(2\pi i x) \)

at exit, \( x=L_x, p=2 \pi U \)

at symmetry plane, \( y=0, \omega=0; z=0, \omega=0 \)

at wall, \( x=0, p=2 \pi U; y=0; \omega=2 \pi \gamma \)

where \( Z, Z_x, Z_y \) are exit, y-wall, x-wall specific acoustic impedance respectively and all are given.

Initial Conditions

As assume the sound field to be caused by the sound source at the entrance, the initial conditions (\( t=0 \)) can be given as:

\[ x=0, p=1; \quad \phi=0, \omega=0; \quad u=v=w=0. \]

(3)
at exit \((i=m)\),
\[
\begin{align*}
\text{for } & \text{plane } y \text{ at } (j=1), \\
& (4p_{i+1} - p_{i-1})/3 \\
& (2u_{i+1} - u_{i-1})/3 \\
& (4w_{i+1} - w_{i-1})/3 \\
& 0
\end{align*}
\]

at symmetry plane \(z \text{ at } (k=1)\),
\[
\begin{align*}
& (4p_{i+1} - p_{i-1})/3 \\
& (4u_{i+1} - u_{i-1})/3 \\
& (4w_{i+1} - w_{i-1})/3 \\
& 0
\end{align*}
\]

\[
\begin{align*}
\text{at exit } & \text{plane } y \text{ at } (j=1), \\
& (4p_{i+1} - p_{i-1})/3 \\
& (2u_{i+1} - u_{i-1})/3 \\
& (4w_{i+1} - w_{i-1})/3 \\
& 0
\end{align*}
\]

NUMERICAL RESULTS

To test the numerical method just presented, a series of sound propagation problems in ducts were calculated. It is shown that the method is successful. Three examples are presented here.

The first example is the sound propagation in hard-wall duct with flow. The initial data are given as follows:

\[
\begin{align*}
\text{M}_{\text{inlet}} &= 0.3, \text{ Z}_{\text{inlet}} = 1, \text{ Z}_{\text{wall}} = \text{Z}_{\text{wall}} - 1. \\
\text{M}_{\text{outlet}} &= 0.3, \text{ Z}_{\text{outlet}} = 1, \text{ Z}_{\text{wall}} = \text{Z}_{\text{wall}} - 1.
\end{align*}
\]

As seen in fig.1, agreement between analytical and the MacCormack method is very good.

![Fig.1 Comparison of pressure profiles](image)

The second example is the 3-D sound propagation in soft-wall duct without flow. Several initial data are given as follows:

\[
\begin{align*}
L &= 1, \text{ Z}_{\text{inlet}} = 1, \text{ M}_{\text{inlet}} = 1, \text{ Z}_{\text{wall}} = 2. \\
\text{M}_{\text{outlet}} &= 0.5, \text{ Z}_{\text{outlet}} = 1, \text{ Z}_{\text{wall}} = 2.
\end{align*}
\]

The results of initial data and some numerical results are shown in table 1, where 4dB is the maximum attenuation of sound energy, \(s\) is the number of time steps at which the equ.(9) is satisfied.

<table>
<thead>
<tr>
<th>(t)</th>
<th>0.0914</th>
<th>0.1827</th>
<th>0.2906</th>
<th>0.587</th>
<th>(\text{dB})</th>
<th>2.997</th>
<th>3.201</th>
<th>2.466</th>
<th>1.784</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.083</td>
<td>-1.040</td>
<td>-0.891</td>
<td>-0.958</td>
<td>0.003</td>
<td>0.0044</td>
<td>0.0087</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0007</td>
<td>0.00007</td>
<td>0.00016</td>
<td>0.00036</td>
<td>0.023</td>
<td>0.022</td>
<td>0.032</td>
<td>0.051</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.018</td>
<td>105.158</td>
<td>102.219</td>
<td>2.018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.018</td>
<td>105.158</td>
<td>102.219</td>
<td>2.018</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The third example is the 3-D sound propagation in hard-soft-hard duct without flow. The initial data are given as follows:

\[
\begin{align*}
L &= 1, \text{ Z}_{\text{inlet}} = 1, \text{ M}_{\text{inlet}} = 1, \text{ Z}_{\text{wall}} = 2. \\
\text{M}_{\text{outlet}} &= 0.3, \text{ Z}_{\text{outlet}} = 1, \text{ Z}_{\text{wall}} = 2.
\end{align*}
\]

Some numerical results are listed as follows:

\[e_2 = \text{0.0056, } 2e_2 = \text{162, } \text{4dB} = \text{1.97}.\]

CONCLUSIONS

The MacCormack method has been used to calculate the 3-D sound propagation in hard-wall and soft-wall ducts with or without flow. The method is more efficient and stable and needs far less storage. Moreover, it is easier to program and debug. It is expected that the MacCormack method would be further applied in sonar propagation.

REFERENCES

PRÉPARGATION EN CONDUITE EN PRÉSENCE D'ÉCOUTEMENT
UNE MÉTHODE TEMPORELLE DE RÉSOLUTION

P. Esposito, Y. Villouvier

Électricité de France D.E.R./Département Acoustique
1, Avenue du Général de Gaulle, 92141 Cléry - FRANCE

INTRODUCTION

Dans les grandes installations industrielles, les sources hydros-électriques ou aéro-électriques de bruit sont souvent localisées à l'intérieur des tuyauteries dans les parties où l'écoulement est fortement perturbé: veines, soupirs, clapets, raccordements de tuyauteries, diaphragmes... Cependant, le rayonnement de cette énergie sonore dans les locaux ou à l'extérieur n'est pas principalement affecté par ces organes eux-mêmes mais plutôt par les très grandes surfaces rayonnantes qui sont les tuyauteries qui leur sont reliées. Or ces conduites sont excitées soit directement par transmission mécanique dans les supports ou les tuyauteries, soit par les ondes sonores qui se sont propagées dans le fluide qui s'y écoulait. Il est donc important dans le cadre des études concernant le bruit engendré par les écoulements de s'intéresser à la propagation sonore dans les tuyauteries soumises à des écoulements quelconques.

Le travail qui est présenté ici concerne une nouvelle méthode de résolution des équations de propagation monodimensionnelle. Cette méthode a été appliquée pour l'instant à l'acoustique en conduite rigide mais écoulement inclue de nombreux développements en cours d'étude.

MÉTHODE DE "PROPAOGATION FAIBLE"

Soit un domaine \( \Omega \) monodimensionnel de variable \( x \), on s'intéresse ici à la résolution de l'un des deux systèmes suivants :

\[
\frac{\partial^2 \mathbf{W}}{\partial x^2} + \mathbf{A} \frac{\partial \mathbf{W}}{\partial x} + \mathbf{B} \mathbf{W} + \mathbf{C} = 0 \quad (1), \text{ ou } \frac{\partial^2 \mathbf{W}}{\partial t^2} + \mathbf{A} \frac{\partial \mathbf{W}}{\partial t} + \mathbf{B} \mathbf{W} + \mathbf{C} = 0 \quad (1)
\]

Dans ces équations \( \mathbf{W} \) (l'inconnue) et \( \mathbf{C} \) sont des vecteurs à deux composantes et \( \mathbf{A} \) et \( \mathbf{B} \) sont des matrices 2x2 dépendant de \( x \). Il s'agit de deux expressions possibles pour un système hybride où le premier ordre linéaire en \( W \).

De tels systèmes s'obtiennent en particulier lorsque l'on écrit l'équation de l'acoustique linéaire dans une conduite monodimensionnelle rigide, en présence d'écoulement.

Soient \( V \), \( \rho \), \( P \), \( C \) et \( S \) (découplés de \( x \) respectivement) la vitesse débitante de l'écoulement, le volume volumique moyen, la pression, la vitesse et la section de la conduite. Ces grandeurs sont des grandeurs caractéristiques de l'écoulement. Soit \( P \) et \( u \) les grandeurs acoustiques qui sont des dérivées d'ordres superieurs de ces grandeurs moyennes. L'une des formulations possibles des équations de l'acoustique en conduite est alors celle de l'équation (1) avec (2):

\[\mathbf{W} = \begin{pmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{B} & \mathbf{C} \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \nabla \mathbf{P} \\ \nabla \mathbf{A} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 \\ \mathbf{C} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 0 \\ \mathbf{C} \end{pmatrix} \]

On peut également mettre sous forme non-conservative ces équations, le système à résoudre est alors le système (1) avec les matrices \( \mathbf{A} \) et \( \mathbf{B} \) légèrement différentes (BxW).

Nous allons décrire la méthode de résolution du système (1), celle du système (2) n'en diffère que par des détails (2) et (1). On cherche donc une solution de (1) sur le domaine \( \Omega = \{ x \} \) et dans l'intervalle de temps \( t^1 \), où \( x \in \mathbb{R}^n \). Si les vecteurs \( \mathbf{P} \) appartiennent à un espace vectoriel \( \mathcal{E} \) muni d'un produit scalaire et si \( \mathcal{E} \) est le dual de cet espace, on peut écrire une formule faible de (1):

\[ \forall \mathbf{y} \in \mathcal{E}, \quad \int_{\Omega} \left( \frac{\partial^2 \mathbf{W}}{\partial x^2} + \mathbf{A} \frac{\partial \mathbf{W}}{\partial x} + \mathbf{B} \mathbf{W} + \mathbf{C} \cdot \mathbf{y} \right) = 0 \quad (2) \]

Les crochets représentent le produit scalaire, c'est-à-dire l'intégrale du produit des fonctions sur \( [a,b] \), \( \in \mathbb{R}^n \).

Dans le formulaire (2), qui est équivalent à la formulation (1), les éléments \( \mathbf{y} \) de \( \mathcal{E} \) sont appelés fonctions-test et dépendent à priori de \( x \) et \( t \).

La formulation (2) est valable pour des champs continus. Si on discrétise les champs, on obtient des espaces vectoriels de dimension finie, qui seront appelés \( \mathcal{E}_h \) et \( \mathcal{E}_h^p \) et on est alors à l'issue d'ax5 (2), que les fonctions \( \mathbf{y} \) appartiennent à une base de l'espace \( \mathcal{E}_h \). On peut choisir une discrétisation de type éléments finis linéaires en utilisant les fonctions scalaire linéaires par morceaux valant 1 au nœud \( a \) et 0 sur tous les autres nœuds. On utilise des "fonctions-chapeau" (ou demi-chapeau a et b) comme celles représentées ci-dessous:

\[ x = \begin{pmatrix} g_1 & g_2 & g_3 & g_4 & \ldots & g_m \end{pmatrix} \]

Ces fonctions permettent de construire 2m fonctions de base et donc l'inconnue \( \mathbf{W} \):

\[ \sum_{i=1}^{2m} \mathbf{w}(x_i) \mathbf{y}_i(x) = \int_{\Omega} \frac{\partial \mathbf{W}}{\partial x} \frac{\partial \mathbf{W}}{\partial x} \mathbf{y} d\Omega \]

Les champs sont donc décrits de manière discrète à partir de 2m valeurs scalaires aux nœuds: \( a \) et de 2m fonctions de base \( \mathbf{y}_i \). Les m premières \( \mathbf{y}_i \) représentent dans notre cas les pressions acoustiques et les m suivantes les vitesses acoustiques. Ces valeurs aux nœuds ne dépendent que du temps tandis que les fonctions de base ne dépendent que de \( x \).

Les méthodes classiques d'éléments finis utilisent pour fonctions-test les fonctions de base que nous venons de décrire. Or ces fonctions ne dépendent que de \( x \) et pas du temps. L'idée de base de la méthode proposée ici (déjà tante à l'application à l'idée de J.P. BENGENE et al. dans le mémoire de conception faible (1)) est l'utilisation des fonctions-test qui ne dépendent que des fonctions de base et qui ne dépendent pas en \( x \).

On part de:

\[ \int_{\Omega} \left( \frac{\partial \mathbf{w}(x)}{\partial t} \mathbf{y}_i(x) + \mathbf{A} \frac{\partial \mathbf{w}(x)}{\partial x} \mathbf{y}_i(x) + \mathbf{B} \mathbf{w}(x) \mathbf{y}_i(x) + \mathbf{C} \cdot \mathbf{y}_i(x) \right) d\Omega = 0 \]

On intègre par parties en temps et en espace et on regroupe les termes obtenus:

\[ \int_{\Omega} \left( \mathbf{w}(x) \mathbf{y}_i(x) dx \right) + \int_{\Omega} \left( \mathbf{w}_i(x) \mathbf{y}(x) dx \right) + \int_{\Omega} \left( \mathbf{w}(x) \mathbf{y}_i(x) dx \right] \]

\[ \int_{\Omega} \left( \frac{\partial \mathbf{w}_i(x)}{\partial x} \mathbf{y}_i(x) - \mathbf{B} \mathbf{w}_i(x) \mathbf{y}_i(x) \right) dx = 0 \quad (3) \]

Ainsi, si l'on sait construire les fonctions \( \mathbf{y} \) de manière à annuler le dernier terme de (3), on simplifie de façon notable l'équation à résoudre. En effet, si on suppose, pour l'instant, que l'on sait trouver ces fonctions, on obtient:

\[ \forall \mathbf{y} \in \Omega, \quad \int_{\Omega} \left( \mathbf{w}(x) \mathbf{y}_i(x) dx + \mathbf{B} \mathbf{w}_i(x) \mathbf{y}_i(x) dx + \mathbf{w}_i(x) \mathbf{y}(x) dx \right) dx = 0 \]

où les termes \( \mathbf{B} \mathbf{w}_i \) représentent les conditions aux limites. Si on fait abstraction de ces termes, qui ne présentent pas de difficultés de calcul (2), et de c qui est nul dans notre cas, le schéma se ramène à:

\[ \forall \mathbf{y} \in \Omega, \quad \int_{\Omega} \left( \mathbf{w}(x) \mathbf{y}_i(x) dx + \mathbf{B} \mathbf{w}_i(x) \mathbf{y}(x) dx \right) dx = 0 \quad (4) \]

Les fonctions-test sont choisies de manière à ce que les fonctions de base à l'instant \( t^{n+1} \) et le système d'équations (4) peut finalement s'écrire sous la forme d'un système de linéaire à 2m fonctions. L'approche la plus naturelle est formée par les produits scalaire de deux à deux des fonctions de base c'est le "matrice de masse" bien connue en éléments finis.
Le second membre de ce système linéaire ne dépend que des coefficients de l'équation (1) et des valeurs des champs à l'instant $t$: $(w_i^e)$. Pour le calculer, il faut construire les fonctions $\psi(x,t)$ sur l'intervalbe de temps $\Delta t$ de manière à obtenir les valeurs de $\psi$ à $t^n$. On doit donc résoudre:

\[ \begin{cases} \psi_i^n t = \frac{\partial \psi_i}{\partial t} + B_j \psi_j = 0 \\
\psi_j(t = t^{n+1}) = \psi_j^n \end{cases} \] (5)

Ce système (5) est résolu à l'aide de la méthode des caractéristiques [2]. Lorsque les matrices $A$ et $B$ sont constantes, la solution obtenue, utilisant les invariantes de Lame, est exacte. Dans les autres cas, il est possible de construire une méthode approchée. La forme générale de la solution est donnée par le schéma suivant, dans lequel on représente les deux composantes d'une fonction-chauffeau particulière aux deux instants $t^n$ et $t^{n+1}$.

Lors du calcul de la forme et de la position des "chauffeaux" de la fonction $\psi$ à l'instant $t^n$, il n'y a aucune raison pour que ces chauffeaux coïncident avec les points du maillage. Il faut donc calculer avec soin les produits scalaires du second membre du système (4). Examinons ce qui se passe pour la première composante de l'un des deux seconds membres. La composante $\psi(x, t^n)$ est représentée par un trait plein tandis que les fonctions de base, invariables, figurent en pointillées.

Parmi toutes les premières composantes des fonctions de base $\Phi_i(x)$, seules celles qui sont "touchées" par $\psi(x, t^n)$ entre les points $A_1$ et $A_2$ auront un produit scalaire non nul avec cette fonction. Il en est de même pour le second composante de $\psi$. Il suffit donc de calculer l'intégrale entre les bornes $A_1$ et $A_2$ en découpant l'intervalle pour n'intégrer que des fonctions quadratiques par morceaux.

PERFORMANCES DE LA MÉTHODE

La méthode, présentée de façon succincte, permet d'obtenir des schémas numériques dans le cas d'une vitesse d'écoulement quelconque et d'un profil géométrique quelconque de la tuyauterie.

Cependant on ne peut effectuer une analyse de l'erreur de l'approximation obtenue que dans le cas particulier où les matrices $A$ et $B$ du système (1) ne dépendent pas de $x$ et lorsque les matelles sont constantes. Cette analyse [2] montre que l'erreur de la méthode est inférieure au millième à partir de 6 points de discrétisation par longueur d'onde. Ces performances sont remarquables, surtout si l'on ajoute qu'il n'y a pas de critère de stabilitéclid au schéma lui-même.

RÉSULTATS

Pour illustrer les performances de la méthode qui vient d'être décrite, on présente la pression et la vitesse de l'onde acoustique le long d'une tuyauterie à un instant donné $t$. L'onde entre ici à gauche dans un écoulement moyen supposé incompressible de 30 m/s dans une conduite de 0,5 cm de diamètre. À 2 m de l'entrée, la section du tuyau est divisée par 10 (donc la vitesse du fluide est multipliée par 10) sur une distance $L$, grande par rapport à la longueur d'onde de l'onde acoustique $\lambda$. La sortie du conduit à droite est anéchique.

On constate que l'énergie acoustique passe facilement à travers le rétrécissement et qu'il n'y a pas d'onde réfléchie. Pression et vitesse acoustiques sont pratiquement en phase partout.

Dans le deuxième exemple, la longueur d'onde $\lambda$ est beaucoup plus grande que la distance $L$ sur laquelle s'effectue le rétrécissement, on peut alors calculer un coefficient de réflexion qui est proche de 0,1 (rapport des surfaces).

CONCLUSION

Sous forme de module de résolution des équations (1) ou (1') avec des conditions aux limites variables, la "propagation faible" permet de résoudre des problèmes simples de propagation d'ondes sonores dans une tuyauterie rigide en présence d'écoulement avec des maillages "assez grossiers".

De nombreux développements sont actuellement en cours d'étude parmi lesquels on peut citer:

- Propagation d'ondes sonores dans les réseaux de tuyauteries.
- Tuyauteries non rigides; couplage avec un calcul mécanique simple de la conduite.
- Passage en dimension 2 ou 3 (pas fractionnaires)

Références


INFLUENCE OF ACOUSTIC WAVE PROPAGATING AXIALLY WITHIN SOUND ABSORBING LAYER IN LINED DUCTS

Zhao Songling and Gong Nongbin
Institute of Acoustics
Tongji University
Shanghai, China

INTRODUCTION

Most porous sound absorbing materials used in lined ducts are usually assumed to be local reacting, and then the acoustic properties of the duct surface depend only on local conditions. In such materials, the acoustic behavior of a point on the boundary surface can be characterized solely by its normal acoustic specific impedance. This approach has the advantage of simplicity since the normal incidence impedance can be easily determined with the aid of impedance tube.

However, as the acoustic wave propagates axially in the tunnel of the lined duct, so does the wave within the sound absorbing layer. In fact, the acoustic behavior of the layer depends on the characteristics of the sound field in the tunnel. Thus, the local reacting assumption is merely applicable to the case in which the influence of the axial wave within the layer can be neglected.

In practice, if the sound absorbing material is rather hard or heavy, the characteristic impedance of the material is much larger than that of the air, and the sound velocity in the material is much lower than that in air, hence, the local reacting assumption is a good approximation. In the opposite case, if the sound absorbing material is very soft or light, as widely used in order to improve its sound absorption property, the normal specific impedance of the sound absorbing layer can be strongly shifted.

In this paper, the influence of the acoustic wave propagating axially within the sound absorbing layer is taken into account. The wave equations as well as the boundary conditions are demonstrated upon rigorous sound propagation theory. Theoretical expressions are derived for rectangular ducts. Typical results are provided and discussed. It is shown that if the sound absorbing material is rather soft or light, the damping coefficient obtained in present theory will be much lower than that predicted in current theory.

Fig.1 Rectangular Lined Duct

In the tunnel, the simplified wave equation is

\[
\left( \frac{1}{c_t^2} + \frac{M}{\bar{C}_L} \right) \frac{\partial^2 p}{\partial t^2} + \frac{\partial p}{\partial x} = 0,
\]

where \( p \) is the sound pressure, \( c_t \) is the sound velocity in static air, \( M=U/c_t \) is the average Mach number in the flow. The symmetrical solution of Eq. (1) is

\[
p = \frac{p_0}{\bar{C}_L} \cos \left( \frac{\pi x}{h} \right) \exp \left( j \omega t - j k_x g x \right).
\]

where \( \omega \) is angular frequency, \( k = \omega/c_t \) is wave number in static air, \( g \) and \( \pi \) are parameters related by the following equation

\[
\frac{\pi}{\sqrt{g}} = (1 - M g)^2 - g^2.
\]

where \( \pi = \omega h/\bar{C}_L \) is the frequency parameter. In the sound absorbing layer, the wave equation becomes

\[
\frac{1}{c_r^2} \frac{\partial^2 p}{\partial t^2} - \Delta p = 0.
\]

where \( c_r \) is the complex sound velocity in the material. Let \( k = \omega/c_r \) be the complex wave number in the material and \( k_y \) be the normal component of \( k \). Considering that the sound wave propagating along \( x \) axis within the layer is similar to that in the tunnel, the form of the solution of Eq. (4) is given by

\[
p = p_0' \cos (k_y (d + h-y)) \exp (j \omega t - j k_x g x).
\]

where \( k_y = k \left( 1 - \frac{c_t^2}{c_r^2} \right)^{1/2} \).

Fundamental Relations

Suppose sound wave propagates along a narrow rectangular lined duct in the presence of air flow as shown schematically in Fig. 1. The halfwidth of the tunnel of the duct is \( h \), and the thickness of the sound absorbing layer is \( d \), the flow velocity in the tunnel is \( U \) in average.

On the boundary surface \( y = \pm h \), sound pressure and normal acoustic displacement (instead of normal velocity) should satisfy the continuity condition[3]. Thus, we get the characteristic equation as follows

\[
\lambda \tan (\lambda k_y) = j (1 - M g)^2 \lambda Y_y,
\]

where \( Y_y = \frac{Y_0 j (k_y/k) \tan (k_y \cdot d)}{\tan (\lambda k_y)} \).

where \( Y_0 \) is the relative value of the characteristic admittance in the material with respect to that in air.

From Eqs. (3), (7), (8), the propagation parameters \( g \) can be solved out for given condition.
Let the imaginary part of \( g \) be \( G \), and the damping coefficient of the lined duct be \( A \) which is defined as the attenuation of sound wave propagated at intervals of \( h \), we have

\[
A = 27.2 \frac{G}{c} \text{ (dB)}.
\]

### RESULTS AND DISCUSSION

In low frequency range, \( c^* \) is much larger than \( |c| \) in practice. From Eq. (6), it can be seen that the normal component of wave number \( k_y \) is approximately equal to the wave number \( k \) in the sound absorbing layer. Hence, the relations mentioned above approach to these obtained under the local reacting assumption. In other words, there is no apparent difference among the present and current theories.

In mediate or high frequency range, the characteristic admittance of the material \( Y_S \) increase with frequency, and the lined duct becomes more effective in general. However, the specific admittance of the sound absorbing layer \( Y \) oscillates around certain intermediate value as frequency increases.

For porous sound absorbing materials, we have the approximate relation

\[
Y_s = \frac{f_c}{f_c \cdot \frac{c}{c^*}}.
\]

where \( f \) is the effective density of the material, \( f_c \) is the density of air. Moreover, the complex factor \( j \cdot \tan(k_y \cdot d) \) oscillates around and tends to unity with frequency, and hence we can obtain the following approximate relation from Eq. (6) and (8)

\[
Y = Y_s \left(1 - Y_0 \cdot g^2\right)^{1/2}.
\]

The above simplified expression is useful for estimating the damping coefficient of the lined duct in the effective frequency range if the band noise is used.

For a given value \( A \), typical contours on \( Y \) plane in static case \( (M=0) \) are plotted in Fig. 2. in which \( \delta = 0.4 \).

It can be seen that if the damping coefficient \( A \) is less than 1 dB, the difference between contours obtained in present theory and Kurze's theory is slight. However, if damping coefficient \( A \) is larger than 2 dB, the difference is quite apparent. In present theory, the damping coefficient \( A \) can hardly be much larger than 2 dB even if the material is very soft.

For given typical values of \( Y_s \), curves of \( A \) versus \( \eta \) are plotted in Fig. 3, in which \( M=0 \), \( \text{Arg}(Y_0) = \pi/8 \) and \( n=1/|Y_0| \).

---

**Fig. 3 Typical Curves of \( A \) Versus \( \eta \)**

It can be seen that if the characteristic admittance \( Y_S \) increase, i.e. the parameter \( n \) decreases, the difference between curves obtained in present theory and Kurze's theory enlarges. For instance, if \( n=1.5 \), the maximum value of \( A \) is 2.4 dB in present theory, and the corresponding value is 3.3 dB in Kurze's theory. However, as frequency becomes sufficiently high so that the frequency parameter is greater than 2, the damping coefficient \( A \) decreases, and the difference of curves obtained in present theory and Kurze's theory reduces.

Considering the effect of air flow, the problem becomes more complicated. In general, to increase the Mach number \( M \) is equivalent to reduce the specific admittance \( Y \) to certain extent \( [4] \), thus, the conclusions mentioned above remain true qualitatively.

Experimental investigation has been performed either. It is shown that experimental results support the present theory.

### REFERENCE

2. Crumer L., Acustica, 3 (1953), 249.
PREDICTIONS AND TESTS OF GAS-DYNAMIC EXHAUST NOISE IN INTERNAL COMBUSTION ENGINES.

G. Ferrari and R. Castelli
Department of Energetics, Polytechnic of Milan; p.L. Da Vinci, 32; 20133 Milano; Italy.

Introduction

Since noise produced by Internal Combustion Engines became more and more important, a section of a comprehensive research program finalized to transport and supported by National Italian Research Foundation CNR, was devoted to the development of new theoretical and experimental techniques to detect and reduce the strength of different noise sources in vehicles.

Within this research program our Institution was charged to set up a suitable approach to the investigation of gas-dynamic noise characteristics generated by intake and exhaust systems of a spark ignition engine. In this paper will be presented the basic assumptions of the model used, the experimental work done to validate computer calculations and, as example of practical solutions proposed, the use of a variable volume Helmotz resonator in parallel with the exhaust duct, to reduce the low frequency noise hardly removed by traditional silencers.

Gas-dynamic noise generation

The gas-dynamic noise radiated from reciprocating engines is due to discharge of exhaust gas from cylinder and to intake of a new charge, at the end of each cycle. Gas flow from cylinder to exhaust duct and from intake pipe to cylinder is clearly unsteady, such as the pressure and velocity wave train which propagates through the ducts to the atmosphere. In Fig. 1 details about this process are given for the particular case of a four stroke exhaust process. When the exhaust valve is opening, the cylinder pressure is about 0.4-0.6 (MPa), while in the exhaust pipe the pressure is about 0.1 (MPa). The pressure ratio through the valve is therefore greater than the critical one and gases are accelerated until they reach the sound velocity in the narrower valve section, from which they expand to the pressure in the pipe. Here gases at the beginning are quiet, but pipe pressure (valve side) is rapidly increasing, because some time is requested to accelerate them and to set up their outflow from the other end. As valve is opening (Fig. 1.a position) pressure head is reducing, because cylinder pressure is progressively decreasing due to gas outflow while pipe pressure is growing, until the gas flow rate from the valve is greater than the outflow from the other end of the duct. When this condition is reversed (Fig. 1.b position) pressure begins to decrease in the pipe too. By this process at the exhaust pipe entrance a pressure wave is generated. It propagates through the exhaust system, where it is partially reflected at any discontinuity and it interferes with other pressure waves by chance met, until it gets last system section from which is partially radiated to the atmosphere. In the same way, the following of induction strokes generates a depression wave trains in the intake system, from whose last section they are transmitted to the atmosphere as sound waves.

Theoretical model

To simulate these processes which appear in the intake and exhaust ducts, the following hypothesis are adopted:
1) Flow is treated as unsteady;
2) Fluid is compressible, without limitation for pressure perturbations;
3) Since longitudinal dimensions are considered prevalent, flow is treated in a one-dimensional manner;
4) Pipe cross-sectional area (A) is variable along the pipe axis with a given law;
5) Process is not incompressible because of friction and heat transfer at the walls.

In our model therefore all quantities are considered constant on a cross-section and functions of the only position coordinate x (along pipe axis) and time t. Assuming as usual, the following quantities are the medium where the pressure p(x,t), the pressure p = ρ(x,t) and the density p = ρ(x,t), the equations of continuity, momentum and energy describing such a flow in a duct have the well-known form:

\[
\frac{dp}{dt} + \rho \frac{du}{dx} + du (\ln \rho) / dx = 0 \quad \text{(continuity)}
\]
\[
\frac{du}{dt} + \frac{1}{\rho} \frac{dp}{dx} + F = 0 \quad \text{(momentum)}
\]
\[
\frac{dp}{dt} - \frac{2}{3} \frac{d}{dt} \left( \frac{1}{\rho} \frac{dp}{dx} \right) = (k - 1) \rho (\lambda + \omega) \quad \text{(energy)}
\]

Many methods of solving these equations numerically have been proposed, but the most widely used because of its simplicity and accuracy is based on characteristics. Unfortunately these calculations are quite tedious, especially when boundary condition is complex. Some attempts have therefore been made to obtain linearized forms of these equations considering just small pressure perturbations in the flow. We preferred the first approach because we were interested not only in acoustic performance of the system, but also in its influence on engine power.

As result of its calculations, our model predicts pressure-time and velocity-time histories in any cross section and particularly at the open end of the exhaust system. Conditions at this boundary are calculated assuming that the flow is quasi-steady and the static pressure remains constant through the boundary and equal to the atmospheric pressure. To compute the radiated noise, it is assumed that the tailpipe outlet acts as simple monopole source \([1,2]\) with a gas flow rate equal to the volume velocity previously calculated. Our case is then equivalent to the problem of the
noise emission by a pulsating body, with its typical solution for sound pressure at a distance 'r' greater than the wavelength [1]:

\[ p(t) = \frac{(\rho \cdot S)}{(4\pi r)} \cdot \hat{u}_0(t - r/a_0) \]

where \( u_0 \) equals medium velocity at the open end of the pipe and \( \rho \cdot S \) equal density and sound velocity in the space surrounding the pipe.

Discussion of results

The accuracy of our model to predict the engine performance and gas-dynamic noise levels was experimentally analysed with a monosylinder 4-stroke engine, supplied with different simple shapes of intake and exhaust system. We used a testing equipment [1] which allowed contemporary power dynamometer tests and noise level measurements in a special test room simulating open field conditions. Choosing suitable empirical coefficients for the model (heat transfer and friction coefficients, reference temperature, etc.) we obtained quite accurate predictions of both noise level and pressure-time histories for some basic systems [1]. In fact, when the model is adjusted by means of an homogeneous set of experimental data, it proves to be quite a suitable tool to optimize the geometry of a given system because it gives the effect at the variation of a single dimension.

As practical solution set up by means of this model, we will give some results we got with a variable volume Helmholtz resonator that we tested in parallel with the exhaust duct to reduce low frequency noise.

![Figure 2](image)

**Fig. 2** - Experimental version of the variable volume Helmholtz resonator tested.

![Figure 3](image)

**Fig. 3** - Pressure values measured (bold line) and computed (dotted line) in the exhaust pipe.

![Figure 4](image)

**Fig. 4** - Measured insertion loss at different engine speeds, produced by the device of Fig. 2.

The geometry of the system was optimised to get best silencing of the low frequency noise (about 100 Hz) without engine performance drawbacks. The resonator was built according to details given in Fig. 2. This solution certainly unsuitable for the practical application to the engine, was considered appropriate for experimental tests because the resonator volume can be easily varied moving by hand step by step the piston. We are now developing a version more suitable for actual use, with the resonator volume coaxial with the duct and an automatic variation of its frequency with the engine speed, through a hydraulic control system.

In Fig. 4 are shown some typical pressure values measured (bold line) and computed (dotted line) in the exhaust pipe (valve side: 68 mm from valve seat) with the resonator excluded or inserted with its optimum volume for the engine speed considered (n = 7000 rpm). These results prove that the model predicts quite accurately the wave action in the system and the resonator modifies the pressure wave trend affecting both engine performance and noise emission. Fig. 4 gives the measured insertion loss obtained at different engine speeds. Also this noise reduction was satisfactorily predicted by the model.

Conclusions

From the available results of the present investigation, the following conclusions can be drawn:

1) Computer predictions provide some useful criteria for a first-attempt design and reduce experimental work, in optimizing intake and exhaust systems;

2) Suitable approach to predict radiated noise is based on modeling the wave action in the manifolds because it provides information about both engine performance and noise emission;

3) A variable volume Helmholtz resonator in parallel with the exhaust duct can effectively reduce the low-frequency noise, hardly removed by traditional silencers of acceptable size.

References


THE USE OF THE CROSS SPECTRAL TECHNIQUE TO THE STUDY OF DISCRETE AND BROAD BAND NOISE

J. S. B. Mather and P. A. C. Medlott

Department of Mechanical Engineering, University of Nottingham, Nottingham. NG7 2RD, U.K.

Introduction

The acoustic output of ducted rotors has been shown experimentally and theoretically by many authors to comprise the summation of the spinning modes of the duct. These researches concerned themselves mainly with discrete energy in the modes and in attempting to predict modal amplitudes and further theories and experimental data followed on many subjects such as the effect of unequal blade signatures on the acoustic output and the generation of modes by distorted inflow. Of late, attention has been moved to random or broad band sources and their relationship to the duct acoustics. The authors have applied the cross spectral technique of measurement and analysis to a wide range of low speed fans, of free and arbitrary vortex design, large and small tip clearances, distorted and undistorted flow, and various blade chord and section combinations. The work is reported extensively in ref. (1). The method has also been applied to high speed precision fans, of the type used in aeroengines, at the National Research Council of Canada, and it is the aim of this paper to extract the essence of both pieces of work.

Cross Spectral Theory

This is well documented (see for instance reference 2) and results in the formulation for the modes of order m, number km as

\[ p^m(a, 0, z; f) = \sum_{m=0}^{\infty} Y_{m} \exp \left(i m \phi + k m \rho \right) \]

where \( \phi = \phi_2 - \phi_1 \) is the circumferential angle between the microphones

\( z = z_2 - z_1 \)

and is the axial distance between microphones for a duct of radius a.

\[ Y_{m} = \sum_{m=0}^{\infty} A_{m} A_{-m}^{*} (J_{m}(k m a))^{2} \]

This assumes that the covariance between the modes is negligible. A spatial Fourier transform of the set of phasors given by this equation will then resolve the sound field into its component modes, both positive and negative. Other modal analysis techniques often only resolve the modes at frequencies which are multiples of the fan speed and thus phase locked to the rotor. The cross spectral method on the other hand gives the modal distribution at all frequencies. Therefore the broad band and narrow band components of fan noise can be resolved into modes as well as the discrete tones. An important rider of this is that in principal, the phase locked components of a tone can be distinguished from the narrow band components by applying both modal analysis methods.

Experimental Rigs

The primary function of the rig was to provide a suitable sound source for the determination of duct modes, in which microphone traverses were incorporated to measure the azimuthal sound pressure pattern without interfering with the aerodynamic and acoustic fields. The high speed fan rig was in one of the anechoic rooms of the Engine Laboratory in the Mechanical Engineering Division of the National Research Council of Canada in Ottawa. It was a single stage axial fan, 30 cm in diameter with 18 rotor blades and 40 stator vanes. The low speed fan rig was sponsored by the Science Research Council in a programme to investigate cooling fan noise. A variety of fans were used, all of approximately 300 mm diameter.

Experimental Results

The determination of the modes by spatial Fourier analysis was accomplished using Hewlett Packard Fast Fourier systems. The converted data was averaged in the frequency domain and the phasors spatially transformed. These dedicated machines reduce the time of modal analysis significantly.

High Speed Fan

Using blade vane ratio theory for the generation of modes by rotating machines with identical blade signatures it is possible to predict the speeds at which modes will cut-on. Table 1 shows these modes.

The power spectra for the four speeds are shown in figure 1. The discrete tone energy at harmonics of blade passing frequency is clearly evident. There is also good agreement with the predictions of Table 1. No discrete energy appears at the first three harmonics at a speed of 3000 rev/min, the first tone being at 3600 Hz, the fourth harmonic. At 4000 rev/min the theory predicts discrete energy at the second harmonic and this is confirmed by the results. Figure 1 also shows that the levels of discrete tones do not necessarily increase with rotor speed. This could be due to a number of reasons, including the effect of the phasing of newly propagating modes of higher radial order than zero, the change in acoustic coupling of modes with the pressure signal generated by the blade, and the relative reduction in the transmission of the signals through the blade row. For the three lower speeds the broad band level rises with speed. At the highest speed this component falls in level perhaps due to the change in aerodynamic efficiency of the unit.

Figure 2 shows examples of the distribution of energy between modes at a given frequency. At both speeds the modes predicted by the theory (Table 1) are confirmed exactly by the results. This agreement was maintained at all of the speeds tested. The cross spectral method as it stands does not of course allow the evaluation of the specific contributions from each of the radial orders associated with each of the circumferential modes. However, the method could be extended to do so using suitably oriented traverses in the radial or axial directions.

There is also some evidence, particularly the amplitude of the second harmonic tone in the spectrum for 4000 rev/min in figure 1, which supports the proposition put forward by a number of authors that the newly cut-on mode (in this case the \((-4,0)\) mode) are more efficient transmitters of acoustic energy. The bulk of the evidence on this subject is contained in the next section.

Low Speed Fan

Tests were carried out on the UK rig on a wide variety of fans, all at low tip speeds (< 70 m/s). The fans were of the isolated rotor type with low aerodynamic performance. They were ideal therefore for studying the propagation of broad band noise which they produced in abundance, as well as for discrete energy. An extensive discussion of the results is given in reference (1) and here we will
restrict ourselves to the major observations and conclusions.

Fig. 3 shows the frequency spectra of the modes for an 8-bladed pressed steel fan. The following observations can be immediately drawn from the figure.

1) Broad band and narrow band noise are transmitted along the duct as both high order spinning modes and plane wave modes.
2) The newly propagating modes cut-on at their theoretical cut-on frequencies.
3) The rise in amplitude of a newly cut-on mode does not necessarily cause a corresponding change in the level of the already propagating modes. This suggests that there is little intermodal covariance.
4) The narrow band of noise indicated by $A_2$ is shown to be made up from two distinct peaks not at the same frequency, each peak corresponding to the positive and negative versions of the same order mode (1,0). This suggests there is little intramodal covariance.

These observations are repeated in all of our test results.

5) The negative modes appear to contribute most of the acoustic energy within the broad band region and the positive mode contribution is mainly restricted to the discrete and narrow band noise.

The first point is well illustrated by figure 4. This figure was generated by collating together the results of several fan tests and normalising the amplitude of the newly cut-on positive mode with respect to the newly cut-on negative mode. For low order modes the graph shows that the majority of points are below the abscissa axis indicating that most of the broad band acoustic energy, regardless of the fan operating conditions, is transmitted by the negatively rotating modes. This might be expected because the axis of the lift force whose fluctuations give rise to a dipole source is orientated in the direction which favours the transmission of negative modes.

The narrow band of noise indicated by $A_2$ in figure 3 is probably the product of an interaction between a rotating distortion and second blade passing frequency. The frequency separation of the positive and negative (1,0) modes is consistent with that which would be expected from an interaction with a distortion rotating at 20 Hz or 0.31 of shaft speed. The greater prominence of a narrow band of noise associated with second blade passing rather than with first blade passing (as shown in $A_1$) suggests that the distortion is predominantly multi-lobed as opposed to single-lobed. The cross spectral modal analysis method would seem a powerful tool for analysing narrow band noise.

Conclusions

The cross spectral technique is able to resolve the sound field into its component modes at all frequencies within the range of interest. It is therefore more useful than the reduction technique although it could be combined with it to resolve narrow band and phase locked components of tones. Doubts over the aspect of inter and intramodal covariances always exist and should be checked in every test situation. It is concluded therefore that

a) discrete energy azimuthal modes can be easily determined,

b) broad band energy is transmitted as spinning modes and plane waves.

c) In general, the most recently cut-on mode, either positively or negatively rotating, transmits more acoustic energy than each of the other

propagating modes. There is not necessarily an equi partition of energy between the modes.

d) The review of the extensive tests on fans suggests that most of the broad band energy, regardless of fan operating conditions, is transmitted by the negatively rotating modes. Conversely, the positive mode contribution is mainly restricted to discrete and narrow band noise.

References

CHARACTERISTICS OF THE FLOW RATE AND NOISE RADIATION OF EXPANSION MICROPOROUS MUFFLER

LI Peizi, MU Xiaomin and WANG Hongyu

Institute of Acoustics, Academia Sinica, Beijing, CHINA

A theory for calculating the flow rate of microporous muffler is proposed. We find a critical expansion ratio below which the flow can only be choked at micropores and noise in the muffler is not important compared with that radiated from micropores. Above the critical, noise produced in expansion effects the noise radiation considerably, particularly at the pressure the flow choking at exhaust pipe only.

INTRODUCTION

The first work of micropore muffler was published in 1977, Ref.1. The noise reduction of a microporous muffler is due to the frequency shift of microjets. The design and manufacturing of microporous mufflers for industrial noise control have been developed and productions are used in blow-offs, particularly in the blow-offs of superheated steam in power station. The noise reduction reaches 40 dB, including the reduction of shock-associated noise.

FLOW CHARACTERISTICS

The micropore muffler is simplified as a combination of three segments of pipes with different cross-sectional area $A_1$, $A_2$ and $A_3$, respectively to exhaust pipe, expansion chamber and all micropores, see Fig.1. We assume that the flow is one-dimension, ideal flow and the stagnation

$$
Q = V P_S M_{i} L A_1 \left[ C_p (\gamma - 1) T_{st} \right]^{-\frac{1}{2}} \times \left[ 1 + \frac{1}{2} (\gamma - 1) M_{i}^{2} \right]^{-\frac{\gamma}{2} (\gamma + 1) / (\gamma - 1)} = \text{const.}
$$

where $\gamma$ is the ratio of specific heats, $C_p$ the specific heat at constant pressure, $P_s$ the stagnation pressure, $T_{st}$ the stagnation temperature, $M_i$ the flow Mach number and 1, 2 and 3, respectively to exhaust pipe, expansion chamber and idealized micropores.

All the numerical calculation taken in the paper are for the air with $\gamma = 1.4$ and $v_o = 1006$ J/kg·K. The reduced equation of conservation of energy is

$$T_{st} = \text{const.}
$$

Equation (1) then can be written in another form

$$M_1 \left[ 1 + \frac{1}{2} (\gamma - 1) M_{i}^{2} \right]^{-\frac{\gamma}{2} (\gamma + 1) / (\gamma - 1)} = \eta M_2 \left[ 1 + \frac{1}{2} (\gamma - 1) M_{i}^{2} \right]^{-\frac{\gamma}{2} (\gamma + 1) / (\gamma - 1)} = \eta M_3 \left[ 1 + \frac{1}{2} (\gamma - 1) M_{i}^{2} \right]^{-\frac{\gamma}{2} (\gamma + 1) / (\gamma - 1)}
$$

where $\eta = \nu_2 / \nu_1$, $\nu_2 / \nu_1$, and $\beta = \nu_2 / \nu_1$, where $\nu_2 = \nu_2 / \nu_1$, and $\nu_2 = \nu_2 / \nu_1$, where $\nu_0$ is the atmosphere pressure. According to the assumption, we have $\nu_2 = \nu_3$, when $M_1 < 1$, the reduced equation of conservation of momentum at expansion is

$$M_3 \left[ 1 + \frac{1}{2} (\gamma - 1) M_{i}^{2} \right]^{-\frac{\gamma}{2} (\gamma + 1) / (\gamma - 1)} = \nu_0 \left[ 1 + \frac{1}{2} (\gamma - 1) M_{i}^{2} \right]^{-\frac{\gamma}{2} (\gamma + 1) / (\gamma - 1)}
$$

Critical $\eta = \eta_C$

At $\eta_C$ flow can be choked at exhaust pipe and micropores simultaneously, i.e. $M_3 = M_3$. From Eq. (3), we have

$$\eta_C = \eta = 1 / x
$$

In Eqs. (3) and (4) the $M_2$ can be eliminated and the relation between $\eta_C$ and is obtained, see Fig.1.

Stagnation Pressure in Micropore Muffler Chamber $\nu_2$ or $\nu_2$

The relation between $\nu_2$ and $M_3$

$$R_s = \nu_0 \left[ 1 + \frac{1}{2} (\gamma - 1) M_{i}^{2} \right]^{-\frac{\gamma}{2} (\gamma + 1) / (\gamma - 1)}
$$

For given $M_1$, the $M_2$ and $M_3$, and $x$ can be obtained from solving the Eqs. (3) and (4), and $R_{s1}$ and $R_{s2}$ then are calculated by Eq. (5).

Flow Rate

The flow rate of microporous muffler can be obtained by substituting the $R_s$ and the total area of micropores into Eq. (1). We assume that the flow rate is ejected from an ideal convergent nozzle at up-stream stagnation pressure $R_{s1}$ and define the exit area of the nozzle the effective area of the muffler, $A_{eff}$. Then we have

$$A_{eff} = \frac{(R_{s2} / R_{s1}) \left[ \frac{\gamma (\gamma - 1)}{(\gamma - 1) / (\gamma - 1)} \right]^{\frac{1}{2}} \times \left( \frac{R_{s1} (\gamma - 1) / (\gamma - 1)}{R_{s2} (\gamma - 1) / (\gamma - 1)} \right)^{\frac{1}{2}} \times \left( \frac{R_{s2} \gamma}{(\gamma - 1) / (\gamma - 1)} \right) \times \eta}{\left( \frac{R_{s1} \gamma}{(\gamma - 1) / (\gamma - 1)} \right) \times \eta \times R_{s1} > \chi \text{ and } R_{s2} < \chi}
$$

where $\chi = \left[ \frac{\gamma (\gamma + 1)}{(\gamma - 1) / (\gamma - 1)} \right]^{\frac{1}{2}}$. The relation between $A_{eff}$ and $A_3 / A_1$ is shown in Fig.2, for $S > 10$ and $R_{s1} > 2$. 

![Graph](image.png)
The marks show experimental results. In the design of a muffler, the $Q$ and $R$, and the cross-sectional area of exhaust pipe $A_e$ are given. $A_{eff}$ can be calculated by Eq. (1). According to that $A_2=0.92 A_1$, see Ref. 2. The $R_e/A_1$ then can be obtained from the theoretical curve given in Fig. 2 by $R_e/A_1 = 0.22$, Ref. 2, the total geometrical area of micropore muffler $A$, to be designed is obtained.

![Fig. 2 Relationship between $A_{eff}/A_1$ and $A_2/A_1$ marks, experimental —— theoretical calculation](image)

NOISE RADIATION

The noise radiation of five micropore mufflers have been studied experimentally and theoretically. The diameters of the exhaust pipe, expansion chamber, and micropores are $d_1=15.6$, $d_2=50$, and $d=1.12$ mm respectively. The five mufflers have different number of micropores being 127, 236, 383, 512 and 639, corresponding to $\eta=0.576$, 0.867, 1.298, 1.735 and 2.165, and $\gamma=10$ and $\eta_a=1.6713$. The experimental average SPLA over a spherical surface with 1 m in radius and centred at the centre of the micropore muffler, $L_a$, are marked in Fig.3. The $L_a$ can be approximated by the following

$$L_a = 80 + 20 \log \left[ \left( R_s - 1 \right)^a \left( R_s - 0.5 \right)^{10} \right] + 20 \log (m \times d)$$

where $R_s$ is the upstream stagnation pressure, $n$ the number of micropores, $x_a=0.165d$, and

$$\Delta L = L_a - L_0 = 24.45 (R_s - 1) \quad \text{when} \quad R_s < 4$$

$$= 6.5 \quad \text{when} \quad R_s > 5 \quad (8)$$

Substituting $R_s=R_1$ calculated into Eq. (7) we obtain the theoretical $L_a$'s of the five mufflers depicted in Fig. 3. It is clear that the experimental results approximate to the calculation if $R_s=1$ and $R_s=2$. Otherwise, the expansion makes noise reduction worse. When the turbulence and noise produced in expansion is reduced that the experimental results approach to the theory is expected, see Fig. 4. The exit area of 295 pores with 1.12 mm pore diameter is equal to that of exhaust pipe with the diameter of 15.8 mm. Therefore, the vertical distance between the upper two curves is the noise reduction due to simple frequency shift of microjets equal to 22 dB. The distance between the lower two curves is the further noise reduction due to

![Fig. 3 Relationship between $L_a$ and $R_s$ for five micropore mufflers with different pore numbers. Experimental results are marked by marks, and theoretical results are marked by curves](image)

![Fig. 4 Reduction of turbulence in expansion](image)

REFERENCES

DESIGN OF AN ANECHOIC CHAMBER TO MEASURE THE SOUND - POWER OF AIR CONDITIONING SYSTEMS.

M. Recuero, C. Gil
Escuela Universitaria de Ingeniería Técnica de Telecomunicación, Complejo Politécnico de Valladolid, Carretera de Valladolid Km. 7, Madrid - 28031, España.

1. Introduction

In view of the importance of Free-field testing and the objections to outdoor arrangements, it is obvious that a Free-field sound room is almost essential for measuring the sound power. The objective in the design of a Free-field sound room is to reduce to a negligible amount all reflections from the boundary surfaces of the room. This is equivalent to a very small ratio of the energy of the reflected sound to that of the incident sound. The ratio of the energy of the reflected sound in a room is

$$\frac{e_r}{e_d} = 16 \times 5^2(1 - \frac{1}{2})/\delta^2$$  (1)

where $e_r$ = energy density of reflected sound, $e_d$ = energy density of direct sound; $\delta$ = distance from the source to the observation point; $S$ = area of absorbing material; $V$ = volume of room; $\delta$ = mean coefficient of sound absorption.

A study of equation (1) shows that the Free-field conditions are approached by making the room large - and absorption coefficient of the wall near unity. To satisfy the first requirement the anechoic chamber must be built as large as possible considering user's needs and architectural and constructional requirements.

The next objective was to obtain an absorption coefficient as near unity as possible. An examination of existing chambers indicates that regardless of the form of treatment, it appears that absorption deviates quite rapidly from unity when the thickness of the treatment is less than a quarter wavelength. In this paper, it is assumed that the effective sound absorption material is measured to an outside boundary is of relatively high acoustical impedance compared with the characteristic acoustic impedance of air.

2. Chamber Acoustic Treatment

During our research work we collected a number of data concerning absorbing structures having different dimensions, in order to find the structure with the simplest construction, lowest cost and optimum performance for any desired frequency range we required.

The structure chosen is a wedge 100 mm long from the base to the apex, its base being a cube with sides of 200 mm. The wedges are made of fiberglass since other possible materials, such as polyurethane, have a high price and do not meet the safety requirements (fire-proof) established by regulations.

The square-based wedge is chosen due to its ease of assembling. The density of the fiberglass chosen on account of its weight, resistance to flow and acoustical impedance is 50 kg/m$^3$. An absorbing element appears in Figure 1. Those wedges will be 4 cm apart - from the concrete wall making a chamber which will be partially filled with a mineral-wool layer fixed to the wall. This structure has an absorption coefficient of 0.99 from 600 Hz up. Figure 2 shows the percentage of reflected energy $q_r$ instead of $q$ as this is a more accurate way of expressing the absorption in highly absorbing structures.

3. Free inner dimensions

According to ISO 3745 - 1977 (8) the volume of the measuring room must be at least 200 times that of the source power level of which is to be measured. Taking into account the fact that the maximum volume of the sources is the equipment of 3 m$^3$ - with a volume of 0.436 m$^3$, the volume of the chamber must be higher than 0.72 m$^3$. In the first calculation we start with a volume of 100 m$^3$ in order to allow the measurement of an equipment of slightly larger volume. Having no specific data, we consider as far field those points located at a distance 2a from the source, where a is the largest dimension in the source (here its diagonal). The measuring surface must be at a distance 3/4 from the absorbing surfaces of the room walls, being $\lambda$ the sound wavelength corresponding to the central frequency of the lowest frequency in which measurements are to be made (125 Hz). The shape of the room will be parallelepipedal.

According to Figure 3 (A and B) the equations proposed in order to obtain the dimensions are the following:

$$\text{height} = 2a + \frac{\lambda}{4} + 0.2 = h$$
$$\text{length} - \text{width} = (2a + \frac{\lambda}{4}) - 2 = s$$
$$\text{volume} \quad V = h \cdot s = 100$$

considering that value 7/4 for 125 Hz is 0.68 m and $\lambda = 0.68$ m, the values for the dimensions are the following: $a = 6$ m, $h = 2.78$ m. The value 0.2 m is the distance between the end of the wedges and the grill which forms the floor.

4. Inside Dimensions

According to the previous data, the inside dimensions of the chamber are:

$$\text{length} = 6 + 2 (0.08 + 0.04) = 7.68$$
$$\text{width} = 6 + 2 (0.08 + 0.04) = 7.68$$
$$\text{height} = 2.78 + 2 (0.08 + 0.04) = 4.46$$

5. Construction of the Chamber

The chamber was constructed of steel-reinforced concrete. The thickness of the side-walls and the floor is 20 cm and that of the ceiling is 10 cm. In order to take the systems to be measured into the chamber easily, the access is at ground level and this requires a hollowing out (Fig. 4). To prevent vibrations coming in, both bottom and sides of the hollow will be covered with a layer of neoprene. The floor of the chamber is a grill platform. The working area is 30 m$^2$. The maximum load will be 300 kg/m$^2$. Table 1 shows the noise level in the area surrounding the chamber. To avoid this noise the construction is double-boxed. Access to chamber is provided by an acoustic door with a system for holding wedge structures so as to preserve continuity of absorbing surface in the room (Fig. 5). The method of installing the wedges is shown in Fig. 6. Each wedge is held in place by screws forming a grid of 20 x 20 cm. A second point of support to prevent sagging of the wedges with age is provided by a new grid 20 cm from the previous one. The air space behind the wedges is filled with mineral wool.

The ventilation system consists of a unit to draw out the air. This is made up of a housing, a double action centrifugal fan, anti-vibration frame, sound-insulating ducts and a flow of 1'600 m$^3$/h.

Fig.1. Wedge, outlines
Fig. 2. Pressure reflection curve

Fig. 3. Free field, (A) plan, (B) elevation

Fig. 4(A). Plan of anechoic chamber

Fig. 4(B). Sectional view of anechoic chamber

Fig. 5. Door of anechoic chamber

Fig. 6. Method of installing wedges

Table 1. Background Noise

<table>
<thead>
<tr>
<th>FREQUENCY Hz</th>
<th>POSITION 1</th>
<th>POSITION 2</th>
<th>POSITION 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>60°6</td>
<td>55°6</td>
<td>51°3</td>
</tr>
<tr>
<td>125</td>
<td>56°6</td>
<td>57°5</td>
<td>50°5</td>
</tr>
<tr>
<td>160</td>
<td>53°6</td>
<td>53°0</td>
<td>50°5</td>
</tr>
<tr>
<td>200</td>
<td>56°3</td>
<td>54°8</td>
<td>50°2</td>
</tr>
<tr>
<td>250</td>
<td>53°3</td>
<td>63°9</td>
<td>50°9</td>
</tr>
<tr>
<td>315</td>
<td>59°9</td>
<td>60°7</td>
<td>51°3</td>
</tr>
<tr>
<td>400</td>
<td>59°6</td>
<td>58°6</td>
<td>53°4</td>
</tr>
<tr>
<td>500</td>
<td>60°7</td>
<td>55°4</td>
<td>58°2</td>
</tr>
<tr>
<td>630</td>
<td>63°0</td>
<td>60°4</td>
<td>58°2</td>
</tr>
<tr>
<td>800</td>
<td>64°2</td>
<td>56°4</td>
<td>54°3</td>
</tr>
<tr>
<td>1000</td>
<td>58°2</td>
<td>60°5</td>
<td>57°4</td>
</tr>
<tr>
<td>1250</td>
<td>62°5</td>
<td>62°9</td>
<td>59°1</td>
</tr>
<tr>
<td>1600</td>
<td>62°4</td>
<td>72°8</td>
<td>53°8</td>
</tr>
<tr>
<td>2000</td>
<td>66°6</td>
<td>57°7</td>
<td>52°4</td>
</tr>
<tr>
<td>2500</td>
<td>59°7</td>
<td>52°1</td>
<td>53°4</td>
</tr>
<tr>
<td>3150</td>
<td>58°6</td>
<td>53°0</td>
<td>49°6</td>
</tr>
<tr>
<td>4000</td>
<td>60°9</td>
<td>53°6</td>
<td>46°1</td>
</tr>
<tr>
<td>5000</td>
<td>53°5</td>
<td>50°9</td>
<td>41°5</td>
</tr>
<tr>
<td>6300</td>
<td>55°6</td>
<td>48°4</td>
<td>39°9</td>
</tr>
<tr>
<td>9000</td>
<td>46°9</td>
<td>40°9</td>
<td>32°5</td>
</tr>
</tbody>
</table>
GUIDELINE FOR REGULATORY CONTROL OF OCCUPATIONAL NOISE EXPOSURE AND HEARING CONSERVATION

S. Gewurtz

Program Development Unit, Ontario Ministry of Labour, 400 University Ave., 9th Floor, Toronto, Canada M7A 1T7

INTRODUCTION

The Federal/Provincial Advisory Committee on Environmental and Occupational Health is composed of senior officials from the health, labour, and environment departments of the federal, ten provincial, and two territorial governments in Canada. It is charged with advising Ministers and Deputy Ministers of Health on all matters of environmental and occupational health. Under the auspices of this Advisory Committee, a working group was established in September, 1982 with specific terms of reference, namely, "to prepare a guideline for regulatory control of occupational noise exposure and hearing conservation with due regard for respective jurisdictions and responsibilities". The membership consisted of two groups. The first was a core group of seven professionals representing Nova Scotia, Ontario, Manitoba, Saskatchewan, Alberta, British Columbia and Health and Welfare Canada, who attended all meetings and who were responsible for specific assignments and for writing the various sections of the final report. In this way, the working group was kept to an economical size for the purpose of attending meetings and preparing documents. The individuals comprising the core group are all active in the field of noise exposure and hearing conservation, and possess expertise in acoustics, occupational health, engineering, industrial hygiene, audiology, physics, and government regulations, with direct access to medical, legal and psychoacoustics expertise as well. The second group comprising the working group were corresponding members from the remaining provincial and territorial governments who were briefed regularly on progress and who could provide input or review documents as appropriate. In this way, the working group maintained a full federal/provincial character.

The working group convened its first meeting in December, 1982 to consider the terms of reference, establish the work programme and allocate assignments and responsibilities. Five additional meetings were subsequently held, the last in November, 1985, to discuss the documents and assignments prepared by members of the core group. Therefore the working group's report represents the consensus of the core group. In addition, an interim report by the working group, dated April, 1984, was circulated to the corresponding members and other interested professionals and organizations. Therefore, the final report which incorporated the comments received has input from the entire working group as well as outside experts.

At the time of writing this paper (January, 1986), the final report has not yet been presented to or approved by the Advisory Committee, and hence it has a draft status at this stage. The final report consists of two parts. Part 1 is the Model Regulation, whose main components are:
- A Framework for a Regulation Respecting Noise Exposure and Hearing Conservation;
- Codes of Practice for Audiology, Hearing Protectors, and Noise Measurement; and
- A Rationale for the Framework.

Part 2 is a COMPLIANCE GUIDE whose components are:
- Guides on Audiology, Hearing Protectors, and Noise Measurement; and
- A Bibliography.

The final report is described in general terms below.

MODEL REGULATION: THE FRAMEWORK AND ITS RATIONALE

In the Model Regulation (Part 1 of the final report), the Framework is structured like a regulation for noise exposure and hearing conservation. Therefore, it provides a guideline which could serve to promote uniformity in such regulations across Canada. However, much as it would be desirable to achieve this objective, it must be recognized that, in Canada, occupational health matters are within provincial jurisdiction, and different agencies have different regulatory approaches and responsibilities. In addition, the resources required to implement the various aspects of a regulation may not always be available in a given agency. Given these considerations, a degree of flexibility must be built into the Framework to allow an agency to meet the requirements appropriately, as required. This flexibility is provided by the Rationale which contains explanatory information on all aspects and every clause of the Framework, as well as alternatives and the factors to be considered in implementation.

APPLICATION OF THE REGULATION

Since the Framework is potentially a regulatory document whose provisions must be clearly understood by all affected parties, both technical and administrative terms used in it must be defined. This is done in the opening section with the proviso that certain administrative terms of a policy nature must be left for an agency to define. One of the key definitions is the one for noise which is an A-weighted sound pressure level equal to or greater than 80 dBA. This definition, when used in section 2 of the Framework, establishes that the regulation applies to every employer and worker at a workplace where any worker is exposed to 80 dBA or greater for a significant period of time during a working day, taking into account the prescribed exposure limits. The application of the regulation is specified in terms of an employer and a worker rather than in terms of noise at a workplace, because both have duties and responsibilities and because this is a health-related regulation directed towards worker exposure. In this regard, two conditions must co-exist in order that the regulation apply, namely:
- noise (i.e. sound levels equal to or greater than 80 dBA) must be present in the workplace; and
- workers are exposed to this noise.

The phrase "for a significant period during a working day" emphasizes these conditions further by specifying that the regulation only applies when the noise to which a worker is exposed exceeds 80 dBA for a time period rather than instantaneously. The phrase "taking into account the prescribed exposure limits" is intended to give some guidance as to what constitutes a "significant period". It should also be noted that noise measurements are not required, but they are recommended when the regulation applies since qualitative criteria exist to indicate whether the sound levels are likely to be above 80 dBA for a significant period of time during a working day.

Since the application of the regulation, as described above, is broad, an agency may wish to specify exemptions. For example, it could be argued that specific industries such as construction and trucking do not lend themselves to provisions which apply to the general industrial establishment. The Framework provides for dispensations to be specified in section 2 at the discretion of the agency.

Closely associated with section 2 regarding the application of the regulation is section 3 which addresses the need for consultation between the employer and workers in the implementation of the regulation. It is by such consultation that workers, through their health and safety representatives or committees, receive information on matters pertinent to the regulation and have input into
the assessment process and the establishment of a hearing conservation program. Such joint participation of employer and worker, sometimes referred to as an internal responsibility system, must exist if the hazard presented by occupational noise exposure is to be addressed effectively.

Assessment Of Noise Exposure

The initial requirement of the regulation is an assessment of noise exposure at the workplace with its inherent positive benefits, and since the application of the regulation is broad, many employers will have to comply with this requirement. As a first step, the assessment can also serve to determine which employers must comply with the more onerous provisions of the regulation. In this regard, there are two conflicting aspects to the assessment. The assessment should be qualitative to avoid the need for mandatory noise measurement in all work places. On the other hand, the results of the assessment should be precise to allow credible decisions to be made.

The novel approach in the Framework is to ascribe these conflicting aspects to two separate components of the assessment. The first is a Screening Assessment of the actual or potential exposure of a worker to noise for a significant period during a work day taking into account the prescribed exposure limits. This is to be accomplished by a medical examination of the work place, not based on noise measurement but taking into account the sources of noise and the nature and extent to which workers are exposed. Clearly, a screening assessment should also be required whenever a change in the work place is made which may alter the noise environment. The conclusions of the screening assessment will indicate whether a more extensive Noise Exposure Assessment, based on noise measurement, is required to determine the noise exposures of workers. This represents the second stage of the assessment process, and is designed to determine if other provisions of the regulation should be implemented. Both the screening assessment and the noise exposure assessment are to be conducted formally, in writing, under employer/worker consultation, and subject to record keeping. However, the screening assessment remains a light burden on employers while providing the benefits of knowledge, awareness and education with respect to noise exposure at the work place.

Hearing Conservation Program

If the noise exposures of workers, measured in the assessment, exceed a specified value and hence constitute an occupational hazard, the employer should take certain measures to protect workers from the noise hazard. The required protective measures constitute a hearing conservation program which can be identified and referred to by all parties.

After careful consideration, the working group recommends as the action level for the hearing conservation program, a time-weighted average exposure (TWA) value of 85 dBA for an 8-hour work day, to be determined with a 3dB exchange rate taking all types of noise into account.

Clearly the hearing conservation program has to be unique to each specific workplace. It may include any, some, or all of the monitoring surveys, controls, hearing protectors, audiometric testing, record keeping, and worker education, as required by other provisions of the regulation.

Noise Exposure Limits And Control

The regulation places a duty on the employer to ensure that the noise exposure of any worker does not exceed specified exposure limits, of which there are two types. The first is a time-weighted average (TWA) exposure limit of 90 dBA for an 8-hour work day. In determining the TWA exposure of a worker, a 3dB exchange rate must be used and all types of noise, including impulse noise must be measured. Since the TWA exposure limit is logarithmic, it serves as a maximum exposure limit for non-impulse noise as well. For example, 90 dBA for 8 hours is equivalent to 114 dBA for 1.5 minutes. However, for impulse noise, the regulation sets a second type of exposure limit, namely a peak sound pressure of 140 dB.

The regulation requires that the primary methods of meeting these limits are engineering controls and work practices. However, where it can be demonstrated that the required controls are not practicable, the employer is permitted to control noise exposure by the use of hearing protectors. In addition, the employer is required to provide a hearing protector upon request by the worker if the worker's TWA exposure exceeds 85 dBA for an 8 hour work day. The regulation places a duty on the worker to use the hearing protector so provided. Both the hearing protectors and their usage must comply with the requirements set out in the code of practice.

Audiometric Tests And Reviews

The employer must provide audiometric tests to all workers covered under a hearing conservation program, and such tests must comply with a code of practice for audiometry. Only the test results determined by the audiometric technician to be abnormal are to be referred to a physician or an audiologist for professional review. If it determined that the abnormal test result is due to noise exposure, both the employer and worker must be advised that remedial action is necessary. Audiometric test results are to be treated as confidential medical information, not accessible by the employer. In addition, the regulation prohibits the employer from taking discriminatory action against a worker whose hearing appears to be affected by noise exposure.

MODEL REGULATION: CODES OF PRACTICE

There are three Codes of Practice included in the model regulation, namely for Audiometry, Hearing Protectors, and Noise Measurement. These codes are short documents and include only mandatory requirements. However an equivalency clause at the beginning of each code allows the employer to vary from the code procedures if the protection afforded to the workers or factors of accuracy are not compromised. Since the codes are referenced in the regulation, they are legally enforceable.

COMPLIANCE GUIDE

A greater degree of flexibility can be realized by dealing with advisory information outside the model regulation. Thus Part 2 of the working group's report contains three guides on Audiometry, Hearing Protectors, and Noise Measurement. These guides are detailed and include recommended practices and descriptive information. They add guidance to the affected parties in the areas addressed by the codes.

CONCLUSION

It is hoped that the model noise regulation, included in the working group report will set the stage for a uniform approach by agencies in Canada for the identification and control of occupational noise exposure. Like any new regulatory initiative, the regulation will be subjected to interpretation. Its provisions will not address all the issues or questions. The resulting dialogue should give rise to a greater appreciation by all concerned parties of the hazards to a worker from noise exposure, and the benefits to be gained from a comprehensive noise regulation which addresses such exposure.
STRADIE RE LA REDUCTION DU BRUIT EN MILIEU DU
TRAVAIL
P. Jousserand**, R. Goula**
* L.A.M.I. 38 rue de la 36 ponte, 31062 Toulouse
Géaux, France
** C.A.G.T. chemin du manel, 31400 Toulouse, France
POURQUOI ET QUAND TRAITER LE BRUIT

La question est d'importance, en effet toute opération consistant à réduire l'exposition du personnel au bruit nécessitera des investissements financiers, une reprise en cause de l'organisation du travail qui ne sont pas parfaitement acceptées "a priori" par la Direction de l'Entreprise ; il importe donc de bien saisir les intérêts de cette démarche pour les individus, pour l'Entreprise et pour la collectivité.

- pour l'individu, il conviendra d'éliminer le risque de surdité professionnelle, affection grave, d'apparition pénible dont on connaît bien hélas les conséquences déplorables. En effet, cette exigence impérieuse, il conviendra d'améliorer au sens large les conditions de travail qui apparaissent de plus en plus comme une exigence légitime des travailleurs. La diminution du stress que le bruit aggrave devra donc être un objectif au même titre que la diminution du risque auditif.
- pour l'Entreprise, le bruit est générateur d'un certain nombre de coûts qui ne sont pas toujours clairement explicités, malfaisants et rebutant, source d'accidents auditifs et non auditifs.
- pour la collectivité enfin, les désordres et les coûts entraînés par le bruit se repèrent sous diverses formes, entraînant un véritable coût social dont les divers éléments commencent à être connus.

Si la connaissance de ces bien favorise la décision d'entreprendre une réduction du bruit, diverses circonstances en faciliteront l'exécution, en sera le cas de la création d'activités nouvelles, de modifications diverses dans l'Entreprise qui constituent un changement, ces situations seront en général plus favorables que celles résultant uniquement de la revendication du personnel ou de l'injonction d'une autorité administrative.

COMMENT TRAITER LE BRUIT

C'est ici que la démarche revêt une importance particulière dont la rigueur méthodologique conditionne en grande partie la qualité des résultats.

La mise en place de la démarche
Elle se fait en clarifiant un certain nombre de points : quelle est l'origine de la démarche, quels sont les objectifs généraux qui l'on souhaite atteindre, quels sont les enjeux qui ont été identifiés, quels sont les acteurs concernés par la démarche et quelle est leur volonté de la voir aboutir.

Le diagnostic de la situation
Il consiste à identifier les problèmes à traiter ainsi que les éléments à préserver, il convient ensuite d'identifier les sources de bruit en les caractérisant, d'identifier enfin les voies de propagation du bruit.

LA DEFINITION DES OBJECTIFS
Il importe qu'elle soit aussi précise que possible et qu'elle fasse apparaître les objectifs liés à la protection de l'ouvr, ceux visant à réduire les perturbations de l'activité de travail, ceux destinés à réduire les effets du stress.

L'ÉTUDE DES SOLUTIONS
Le panorama des différentes solutions envisageables est à examiner dans son ensemble, à ce stade, les actions de réduction du bruit pouvant porter sur les sources, sur la propagation, au niveau de la réception du bruit ainsi que sur l'isolement des locaux.

Les méthodes de l'acoustique prévisionnelle sont ici d'un grand intérêt pour tester les différentes solutions par rapport aux objectifs précédemment définis.

Les choix
Les peuvent alors être effectués en toute connaissance de cause, avantages, contraintes, coûts ayant été correctement appréciés au cours des étapes précédentes.

La réalisation et sa vérification
La réalisation s'appuiera sur un cahier des charges bien élaboré, elle sera l'objet d'un suivi attentif et les vérifications finales de résultats seront facilitées par la présence du cahier des charges.

AVEC QUI TRAITER
Tout au long de la démarche, il est absolument indispensable d'associer aux différentes étapes tous ceux qui sont concernés par le problème du bruit, directement comme les opérateurs exposés, indirectement comme ceux se portant à divers services de l'Entreprise en particulier, bureau des méthodes....

Pour atteindre ces objectifs, il importe que la réduction du bruit soit le fait d'une démarche globale, pluridisciplinaire dans laquelle s'impliquent aussi bien les acteurs appartenant à l'Entreprise que les partenaires extérieurs dont elle se sert pour réaliser des études, conseils, entreprises de traitement....

C'est aujourd'hui la condition indispensable pour que la volonté de réduction du bruit en milieu du travail ne se traduise pas uniquement en nombre de décibels gagnés mais débouchera sur une véritable amélioration des conditions de travail correspondant au mieux aux attentes des différents acteurs de la vie dans l'Entreprise.

REFERENCES
THE RELIABILITY OF PERSONAL NOISE DOSEMETERS UNDER STEADY-STATE AND VARIABLE NOISE EXPOSURE

M. Rheault and R. Héroux

GAUM, Université de Montréal, Pavillon Marguerite d'Youville, C.P. 6128, succursale A, Montréal, Québec, Canada, H3C 3J7

The widespread use of personal dosimeters have been justified by the need to measure directly the exposures to fluctuating noise, and more especially for noise fluctuations that depend on motions of the individuals in different worksites. The meaningfulness of the measure in terms of risk of hearing loss is a function of its reliability: it must represent an accurate estimate of the long-term exposure (1). Sources of error have been identified for personal exposure meters: the microphone location (2-3), and its interaction with the nature of the sound field (4-5), the accuracy of the frequency response complying to Type 2 sound level meter tolerances limits (6), the limited dynamic range of the device (7) and its response to high level impulses (8-10). The variability of the measure to be measured has been considered (11) but few studies have attempted to quantify the actual contribution of this source of error.

The aim of the present investigation was to assess the reliability of personal dosimetry for different patterns of exposure in industry. Three categories of exposures were defined for the purpose of this study: - category 1 (C1): job assignments that involve constant daily exposure to steady-state noise without motion in space; - category 2 (C2): job assignments involving displacements in different noisy areas that are predictable in space and time, or exposure to time-varying noise, the variations being determined and predictable for a workday and repeated from one day to another; - category 3 (C3): job assignments that are partially or totally unpredictable, involving varying exposures within a workday and from one day to another.

It was hypothesized that the dose measurements over a workday would be highly reliable for the first two categories and relatively unreliable for the third one.

METHODS

Selection of the industrial settings

In order to minimize the possible influence of other sources of variation, the plants selected had to meet the following criteria: absence of audible discrete impact/impulse noises or predominantly high frequency noise (above 3 kHz). A weaving mill and a food processing plant were found to meet these criteria.

Subjects

For each of the three categories of exposure, a group of 15 volunteer workers was selected. Their job assignment and pattern of noise exposure had to fit the definitions given above. Subjects in group C1 were production workers. Group C2 comprised production workers and maintenance personnel assigned to a restricted area in the production line. Workers belonging to group C3 were responsible for the general maintenance in the factory.

Equipment

Ten DuPont MX1 dosimeters were used. They were submitted to a thorough verification and calibration prior to the experiment. The calibration was checked before and after every day of measurement. They were set to operate with a 2 dB exchange rate and a threshold of 80 dB(A). A Brüel & Kjaer integrating sound level meter (model 2223) was also used for a parallel assessment of the exposures of the 45 subjects.

Procedure

Each subject wore the same dosimeter for a full 8-hour work shift during three consecutive days. They received instructions to prevent artifact in the measurements. The microphone, attached to the clothing, was located at the shoulder.

Exposure measurements with the sound level meter involved the following steps: the worker was first met to obtain a detailed description of his work organization and schedule, then measures of noise were obtained for each activity or job performed during a representative workday. For the subjects belonging to group C3, a list of assignments within a "typical" week or month period was first obtained; the sampling of the noise exposure levels was then performed accordingly.

RESULTS

Table 1. Mean, standard deviations (s), minimum and maximum dose values during 8-hour periods) for three categories of exposures. The corresponding results obtained with the sound level meter (SLM) are also given.

<table>
<thead>
<tr>
<th>Group</th>
<th>Category</th>
<th>Mean</th>
<th>s</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>day 1</td>
<td>94.96</td>
<td>6.58</td>
<td>86.4</td>
<td>106.4</td>
</tr>
<tr>
<td>C1</td>
<td>day 2</td>
<td>94.50</td>
<td>6.41</td>
<td>85.9</td>
<td>105.9</td>
</tr>
<tr>
<td>C1</td>
<td>day 3</td>
<td>94.25</td>
<td>6.13</td>
<td>86.8</td>
<td>103.9</td>
</tr>
<tr>
<td>C1</td>
<td>SLM</td>
<td>94.15</td>
<td>6.47</td>
<td>85.9</td>
<td>103.1</td>
</tr>
<tr>
<td>C2</td>
<td>day 1</td>
<td>87.92</td>
<td>7.47</td>
<td>77.2</td>
<td>101.5</td>
</tr>
<tr>
<td>C2</td>
<td>day 2</td>
<td>88.01</td>
<td>7.31</td>
<td>73.5</td>
<td>101.6</td>
</tr>
<tr>
<td>C2</td>
<td>day 3</td>
<td>88.49</td>
<td>6.44</td>
<td>81.0</td>
<td>101.2</td>
</tr>
<tr>
<td>C2</td>
<td>SLM</td>
<td>88.08</td>
<td>6.68</td>
<td>79.1</td>
<td>101.2</td>
</tr>
<tr>
<td>C3</td>
<td>day 1</td>
<td>85.78</td>
<td>6.06</td>
<td>74.4</td>
<td>98.1</td>
</tr>
<tr>
<td>C3</td>
<td>day 2</td>
<td>84.32</td>
<td>8.52</td>
<td>69.7</td>
<td>100.0</td>
</tr>
<tr>
<td>C3</td>
<td>day 3</td>
<td>86.28</td>
<td>7.04</td>
<td>71.9</td>
<td>97.4</td>
</tr>
<tr>
<td>C3</td>
<td>SLM</td>
<td>89.32</td>
<td>4.26</td>
<td>82.9</td>
<td>99.2</td>
</tr>
</tbody>
</table>

The mean results presented in Table 1 are in agreement with our assumption: for exposures to steady-state noise (C1) and to predictable fluctuating noise (C2), the average dosimeter readings for a group were highly reproducible from one to another and they were very similar to those obtained with a SLM. For relatively unpredictable time-varying noise (C3), the daily mean doses showed variations; the standard deviations are also more variable and tend to be higher than in the other groups. The range of daily dose extend from much lower minimum values to approximately the same maximum values. Moreover, there is a clear disagreement between the dosimetric and SLM results. These observations are confirmed by the results of the analysis of variance (randomized block design) on the factors "repetition of measurements" and "method of measurement" as shown in Table 2.

The results from Table 2 demonstrate that there was no systematic variation in daily noise dose measurements. But one will note that the variability (in dB squared) was much higher for group C3. When comparing SLM results with the logarithmic average of the dosimetric results, significant differences are obtained for categories C1 and C3. In the first case, the mean difference is equal to 0.5 dB in favor of the dosimetric readings. This is explained by a very slight but systematic overestimation of the time spent away from the noise (e.g., in the lunchroom) when interviewing the workers about their work schedule. This small bias was probably present for the
Table 2. Results of one-way analysis of variance on the effect of the repetition of measurements with dosimeters and method of measurement, for each category of exposure.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Repetition of meas.</th>
<th>Method of meas.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>F</td>
</tr>
<tr>
<td>Group C1</td>
<td>1.95</td>
<td>1.63</td>
</tr>
<tr>
<td>Group C2</td>
<td>1.38</td>
<td>0.51</td>
</tr>
<tr>
<td>Group C3</td>
<td>1.56</td>
<td>1.41</td>
</tr>
<tr>
<td>Group C1</td>
<td>1.67</td>
<td>4.96</td>
</tr>
<tr>
<td>Group C2</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Group C3</td>
<td>8.41</td>
<td>8.40</td>
</tr>
</tbody>
</table>

Other categories, but it would have been outweighted by the higher intra-individual variability in the exposure along the day. The effect of the mode of measurement for category C1 is explained by the fact that with the SLM method, the exposures were estimated over 40-hour and 160-hour periods and then converted to daily 8-hour doses. The limited 1-day dosimetric sampling did not take into account some of the most severe exposures occurring over a typical week or month interval.

In the absence of any systematic daily variation in dosimetric results, the random error was further analysed by computing the standard error of measurement: it is based on the reliability coefficient, as indicated in the following equation (12):

\[ S_e = S(1 - R_{xx})^\frac{1}{2} \]

where

- \( S \) = standard deviation
- \( S_e \) = standard error
- \( R_{xx} \) = reliability coefficient

Table 3. Reliability coefficients, standard error and 95% confidence intervals of dosimetric measurements in dB for the three categories of exposures.

<table>
<thead>
<tr>
<th>Group</th>
<th>( R_{xx} )</th>
<th>( S_e ) (dB)</th>
<th>( \pm ) 1.96 ( S_e ) (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.97</td>
<td>1.10</td>
<td>( \pm ) 2.16</td>
</tr>
<tr>
<td>C2</td>
<td>0.95</td>
<td>1.56</td>
<td>( \pm ) 3.06</td>
</tr>
<tr>
<td>C3</td>
<td>0.83</td>
<td>2.86</td>
<td>( \pm ) 5.61</td>
</tr>
</tbody>
</table>

Assuming that the measurement error is independent of the magnitude of the measure and that it is normally distributed, the values given in Table 3 represent estimates of the variability of individual results; this allows to define confidence intervals of individual exposure levels measured by means of personal dosimeter. Thus, for 95% of the cases in group C1, the results obtained is within \( \pm 2.2 \) dB of the true dose. For group C2 and C3, the error margin extends over 6.2 and 11.2 dB respectively. It implies for example that a dosimetric result of 90 dB indicates that the true exposure level is somewhere between 87.8 and 91.2 dB in group C1, between 86.9 and 91.2 dB for a worker in group C2 and between 86.6 and 91.6 dB for an individual belonging to group C3.

**DISCUSSION**

Our hypothesis is confirmed by the present results obtained with group C3: unless it is conducted on a homogeneous group of workers, personal dosimetry is relatively unreliable when evaluating daily exposures that are partially or totally unpredictable. An appropriate identification of the exposure variables is necessary to accurately assess representative daily doses. Repeating the measurement over three consecutive days was not sufficient to achieve this accurate estimation; averaging the doses over three days of measurement would only reduce the variability by a factor of 1.7 (that is the square root of 3). The error margin of the average would be \( \pm 3.2 \) dB, a range of values that cannot be considered as negligible. Consequently, at least for this category of exposures, a systematic analysis of the work organization within the appropriate time scale combined with an adequate sampling of the noise levels with a precision SLM is certainly more reliable. Furthermore, it was less time consuming to perform direct measurements at several sites and for several activities than repeating personal dosimetry over three days (which were insufficient to achieve a representative estimate of the individuals long-term exposure levels).

However, the reliability of the SLM survey method also be assessed for this type of unpredictable exposures; independent estimates of the long-term exposures of general maintenance personnel may turn out to be relatively variable, unless considerable time is devoted to the survey of the exposure of each worker.

The results obtained with exposure categories C1 and C2 did not confirm our hypothesis at the level of individual measurements. Despite a careful selection of the workers in accordance with the definition of our exposure categories, only group estimates can be said to be reliable. The individual readings are subject to a significant daily variation; even if they are averaged over three days of measurement, their error margin is still significant: \( \pm 4.2 \) dB for group C1 and \( \pm 4.8 \) dB for group C2. Considering that, using the SLM, it takes only a few minutes to obtain several measurements of exposure of a worker belonging to group C1, it certainly represents a more valid and practical method of dose assessment. But it is also the case for workers belonging to group C2, even though the sampling of the noise levels for different activities require more time; attaining a higher degree of reliability can be achieved in less survey time using an integrating sound level meter and analysing the work organization than undergoing personal dosimetry over several days with the same worker.

Considering the influence of the variability of exposure and the other sources of error of personal dosimetry (1-10), one can conclude that this approach to noise exposure measurement is of limited use in industrial settings (11).

**REFERENCES**

UN MODELE D’ÉVALUATION ÉCONOMIQUE DES ACTIONS DE PRÉVENTION CONTRE LE BRUIT.

T. Schneider

INSERM U240, BP 48, 92263 Fontenay-aux-Roses CEDEX.

En France, les surdités liées à l’exposition au bruit sur les lieux de travail sont reconnues comme maladies professionnelles ouvrant droit à réparation, depuis 1963. En 1981, un décret a modifié sensiblement les conditions de reconnaissance et de réparation associées à ces surdités, et de l’augmentation considérable du nombre de surdités professionnelles indemnisées avec près de 1 300 nouveaux cas en 1983, soit deux fois plus qu’en 1981 l’1/1. Cette nouvelle situation passe avec encore plus d’acuité le problème de la prévention contre le bruit en milieu industriel. Il paraît probable que si les entreprises ne réalisent pas, dès à present, des investissements permettant de réduire les niveaux sonores sur les lieux de travail, le nombre de nouveaux cas de surdités professionnelles reconnues chaque année prendra à accroître, entraînant à la fois une augmentation des coûts de réparation supportés par les entreprises, ainsi qu’une augmentation des coûts sociaux relatifs à ces surdités. Cependant, au-delà des problèmes techniques qui ne sont pas négligeables, l’amélioration de la protection apparente, bien souvent, pour les entreprises comme une charge financière importante et dissuasive. En ne considérant que les effets auditifs liés à l’exposition à des niveaux sonores élevés, il est important, pour incliner la mise en œuvre de politiques de prévention, de pouvoir quantifier les risques potentiels de surdités professionnelles pour des populations restant exposées à des niveaux sonores encore élevés, et d’évaluer les efforts à attendre d’une réduction des niveaux sonores. L’analyse économique consiste alors à déterminer les coûts de réparation associés à ces effets et à les comparer aux coûts relatifs à la mise en œuvre de la prévention.

ÉVALUATION DES RISQUES DE SURDITÉS PROFESSIONNELLES

Plusieurs modèles ont tenté de mettre en relation l’exposition des travailleurs à des niveaux sonores élevés et les déficits auditifs induits par le bruit. Cependant, le modèle le plus approprié à l’évaluation des risques de surdités professionnelles est basé sur la condition de reconnaissance du système français, est le modèle présenté dans la norme AFNOR (NF S 31 013) /2 reprenant les résultats de la norme internationale (ISO/DIS 1999). Il permet d’estimer des déficits auditifs permanents liés à l’âge et au bruit pour une population donnée. En fonction de la durée d’exposition, de l’âge et du sexe. Il est cependant indispensable d’en noter ses limites : les niveaux sonores à considérer doivent être compris entre 85 dB(A) et 100 dB(A) et, de même, les durées d’exposition doivent être inférieures à 40 ans et l’âge des populations inférieur à 60 ans (les bases de données ayant permis l’élaboration de cette relation étant comprises dans ces limites). De plus, s’agissant ici des données statistiques, l’évaluation du déficit auditif d’individus pris isolément ne peut être établie. De même pour les quantités de population inférieurs à 5% ou supérieurs à 95% les données expérimentales ne sont pas suffisantes pour valider ces intervalles. Dans la nouvelle norme française, l’estimation des déficits auditifs liés à l’âge et au bruit d’une population notée \( D_{O} \) (indice Q représentant le quantile de la population considérée) s’effectue par la relation suivante :

\[
D_{O} = A_{O} + B_{D}Q - (A_{Q} - B_{D}Q)(1/20),
\]

où \( A_{O} \) est le déficit auditif lié à l’âge pour le quantile \( Q \) de la population, et \( B_{D} \) est le déficit auditif, induit par le bruit.

Les relations précédentes permettent de calculer les déficits auditifs. Cependant, par souci de simplification, le calcul n’a pas été effectué pour tous les âges et toutes les durées d’exposition mais pour des tranches d’âge et d’exposition de cinq années. De même, seuls les niveaux sonores de 85, 90, 95 et 100 dB(A) ont été retenus. En tant que l’évaluation des déficits relatifs à une population exposée n’a de sens que statistiquement, cette simplification n’altère pas la signification des résultats. Neuf classes d’âge sont constituées entre 20 et 60 ans ainsi qu’une durée d’exposition entre 5 et 40 ans. Le croisement de ces deux variables conduit à définir quarante-quatre cohortes possibles pour les hommes et autant pour les femmes. Seules les cohortes pour lesquelles le risque d’apparition d’un déficit auditif supérieur à 30 débits dépasse 5% sont retenues. Le déficit auditif est calculé à partir de la définition donnée par la Sécurité Sociale pour la reconnaissance d’une surdité professionnelle faisant intervenir les déficits pour les quatre fréquences suivantes : 500, 1 000, 2 000 et 4 000 Hertz. La formule suivante permet de calculer ce déficit :

\[
D = (2D(500) + 4D(1 000) + 3D(2 000) + D(4 000))/(10)
\]

où \( D(500) \) (respectivement 1 000, 2 000 et 4 000 Hz) est le déficit auditif mesuré à la fréquence 500 Hz (respectivement 1 000 Hz, 2 000 Hz et 4 000 Hz) \(/\). Afin de bien les tableaux de risques, trois intervalles de déficit auditif ont été retenus pour lesquels il est possible, à partir du bas niveau indicatif de la Sécurité Sociale, de faire correspondre des coûts de réparation différents (il s’agit de déficits auditifs compris entre 30 et 35 dB, entre 35 et 43 dB et entre 43 et 55 dB). Pour les hommes, quatre tableaux de risques sont obtenus correspondant à des explications aux niveaux sonores 85, 90 et 100 dB(A). Par contre, pour les femmes, seule l’exposition aux niveaux sonores 95 et 100 dB(A) entraîne des risques de surdité supérieurs à 5%, ainsi deux tableaux sont obtenus. La différence de déficit entre les cohortes masculines et féminines est expliquée par le fait que le déficit lié à l’âge est toujours moins important chez les femmes. Le tableau 1 présente quelques risques de surdité en fonction des trois classes de déficit auditif \(/\).

Tableau 1 : Risque (en %) pour le niveau sonore 100 dB(A) pour une population masculine âgée de 35 ans.

<table>
<thead>
<tr>
<th>DURÉE D'EXPOSITION (en années)</th>
<th>DÉFICIT AUDITIF (en dB)</th>
<th>50-55</th>
<th>35-45</th>
<th>45-55</th>
<th>&lt; 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>9</td>
<td>7</td>
<td>0</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>15</td>
<td>0</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>11</td>
<td>12</td>
<td>0</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>11</td>
<td>14</td>
<td>5</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>11</td>
<td>15</td>
<td>6</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>12</td>
<td>15</td>
<td>7</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

L’intérêt de tels tableaux est de permettre la détermination, non pas de risques individuels, c’est-à-dire la probabilité d’apparition d’une surdité pour un individu particulier, mais de risques collectifs, c’est-à-dire le nombre attendu de surdités pour une population donnée. La correspondance entre les caractéristiques des individus et les tableaux de risques fournit une estimation concernant les risques actuels. Une deuxième estimation, plus intéressante dans l’optique de la prévention, peut être obtenue en déterminant les risques d’apparition de surdités professionnelles dans plusieurs années (la période retenue est de 5 ans). Les caractéristiques de la population sont obtenues en ajoutant cette période aux âges et durées d’exposition initiaux. L’hypothèse consiste à considérer que l’ensemble de la population sera encore exposé durant cette période. Le turn-over n’est toutefois pas introduit dans ces calculs. Cependant, il est nécessaire de ne plus considérer l’évolution des risques pour les individus ayant atteint l’âge de la retraite. Enfin, une troisième estimation consiste à déterminer les risques d’apparition des surdités professionnelles dans plusieurs années mais en considérant que la réalisation actuelle d’une action de préven
tion réduit le niveau sonore équivalent auquel est exposée la population. Pour cette estimation, il est nécessaire, dans un premier temps, de calculer deux catégories de risques : les risques de délicat auditif encourus par la population dans plusieurs années d'une part, sans aucun changement du niveau sonore et, d'autre part, en supposant que le niveau sonore a toujours été celui obtenu suite à la réalisation de la prévention. Ainsi, en considérant la réalisation actuelle d'une action de prévention, les risques d'apparition des surdités professionnelles dans plusieurs années sont calculés en pondérant ces deux catégories de risques par les durées effectives d'exposition à chacun des niveaux sonores retenus. Pour chaque individu ou groupe d'individus ayant des caractéristiques semblables, le risque de déficit auditif, $R$, est défini par la relation suivante :

$$R = R_1 \times (d/D) + R_2 \times (X/D)$$

où $R_1$ est le risque de déficit auditif avec le niveau sonore initial, au bout d'une période d'exposition égale à $D$ ; $R_2$ est le risque de déficit auditif avec le niveau sonore obtenu après correction, au bout d'une période d'exposition égale à $D$ ; $d$ est la durée d'exposition actuelle de l'individu au niveau sonore initial ; $X$ est le nombre d'années d'exposition suite à la réalisation de la prévention (ici $X = 5$) ; $D = d + X$, est donc la durée totale d'exposition.

Cette relation permet de constituer des tableaux traduisant les risques d'apparition des surdités professionnelles cinq ans après la réduction du niveau sonore équivalent. Le graphique I présente l'évolution, après cinq années, du risque d'apparition de surdité (c'est-à-dire pour un déficit supérieur à 30 dB sans distinction entre les trois intervalles) pour une population masculine exposée à 100 dB(A) : la première courbe présente le risque si aucune prévention n'est réalisée, tandis que la seconde présente le risque après une réduction du niveau sonore à 85 dB (A).

Graphique I : Courbes de risque d'apparition de surdité professionnelle dans cinq ans.

<table>
<thead>
<tr>
<th>DURÉE EXPOSITION</th>
<th>RISQUE CALCULÉ</th>
<th>RISQUE RÉALISÉ</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>1</td>
<td>0,25</td>
</tr>
<tr>
<td>30</td>
<td>1,5</td>
<td>0,30</td>
</tr>
<tr>
<td>35</td>
<td>3</td>
<td>0,45</td>
</tr>
<tr>
<td>40</td>
<td>4</td>
<td>0,60</td>
</tr>
<tr>
<td>45</td>
<td>5</td>
<td>0,75</td>
</tr>
<tr>
<td>50</td>
<td>5,5</td>
<td>0,90</td>
</tr>
<tr>
<td>60</td>
<td>7</td>
<td>1,20</td>
</tr>
</tbody>
</table>

ANALYSE ÉCONOMIQUE

L'intérêt de l'évaluation des risques d'apparition de surdités professionnelles parmi une population exposée est d'avoir un aperçu des industries de l'ère de manière prévisionnelle. Une surdité reconnue comme maladie professionnelle, entraîne des coûts de réparation pour l'entreprise (J). Dès lors, l'analyse économique des actions de prévention doit tenir compte d'une part, des coûts relatifs à la réalisation de l'action de prévention et d'autre part, les coûts de réparation évités du fait de la diminution des risques suite à la réduction du niveau sonore. Les coûts consécutifs à la mise en œuvre d'une action de prévention intégreront généralement deux composantes : les dépenses d'investissement associées à l'installation de la mesure et les dépenses d'exploitation (entretien, énergie…) qui sont évaluées sur une base annuelle. Les coûts de réparation sont quant à eux calculés à partir des cotisations versées à la Sécurité Sociale pour la réparation des surdités professionnelles. Ces dernières sont fonction de la rente d'incapacité permanente associée à la surdité. Pour le calcul de ces coûts de réparation il est fait l'hypothèse que l'ensemble des risques est supporté cinq années après la réalisation de l'action de prévention. Pour l'entreprise, ces dépenses futures ne peuvent être comptabilisées au même titre que les coûts de prévention étant donné que, du point de vue monétaire, une dépense actuelle n'a pas le même poids qu'une dépense future. Afin d'intégrer ce comportement économique de l'entreprise, les coûts de réparation sont donc actualisés. La figure 1 présente le principe de l'analyse économique.

![Figure 1 : Analyse économique](image)

Le bilan économique d'une action de prévention est obtenu en comparant les coûts de réparation évités aux coûts de prévention. L'exemple suivant a pour objectif de montrer la faisabilité d'une telle évaluation mais ne constitue pas une analyse des solutions acoustiques. Il s'agit du capotage d'une machine dans un atelier concernant quinze ouvriers, lequel permet de réduire le niveau sonore dans l'atelier de 100 dB(A) à 85 dB(A). Le tableau II présente les résultats obtenus.

<table>
<thead>
<tr>
<th>Déficit auditif</th>
<th>SANS PREVENTION</th>
<th>AVEC PREVENTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-35</td>
<td>1,16</td>
<td>1,10</td>
</tr>
<tr>
<td>35-45</td>
<td>0,99</td>
<td>0,70</td>
</tr>
<tr>
<td>45-55</td>
<td>0,85</td>
<td>0,45</td>
</tr>
<tr>
<td>&gt; 50</td>
<td>0,70</td>
<td>0,40</td>
</tr>
</tbody>
</table>

Ainsi, les coûts de réparation évités actualisés à 9 % s'élèvent à 120 000 F., alors que la réalisation du capotage entraîne une dépense de 70 000 F. Le bilan de l'action de prévention se traduit par une diminution de 0,85 des risques et une diminution des dépenses de 50 000 F.

CONCLUSION

L'application du modèle à l'évaluation de plusieurs actions de prévention contre le bruit en milieu industriel a démontré son caractère opérationnel. Cependant, dans l'état actuel de son développement, il ne s'agit que d'une première étape dans la mesure où le modèle ne prend pas encore en compte les effets extra-auditifs associés au bruit.

REFERENCES

REHEARSAL STUDIO ACOUSTICS AND THE SOUND EXPOSURE EXPERIENCED BY MILITARY BANDSMEN

R.B. Crabtree

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West
P.O. Box 2000
Downsview, Canada, M3M 3B9

BACKGROUND

Military bands have undergone a period of transition that has seen the traditional instruments of marching ensembles augmented by those more closely associated with stage performance. Many modern concert bands now feature full tympani and percussion sections. Amplified instruments such as guitars, basses, pianos, organs and synthesizers have become important parts of such aggregations, and sound re-enforcement systems are commonplace for vocal and solo instrument work. Repertoires are keeping pace with modern times, with a proportion of "high energy" music that fully exploits the capabilities of the newer instruments. Hence, there is a possibility that the members of these bands will incur a greater sound exposure than they would otherwise.

In all, there are nine concert bands in the Canadian Forces (CF). Some of the members of these bands also play in smaller associated ensembles such as marching or dance bands and combos, as required. By tradition, full-band rehearsals take place in the mornings, and individual practice or related activity, in the afternoons. Since considerable rehearsal time can be spent in preparation for a performance, the greatest exposure to sound would seem to occur during these sessions.

Requests for the participation of the Defence and Civil Institute of Environmental Medicine in the studies of CF band facilities were precipitated both by anecdotal evidence from musicians that exposure to sound during their rehearsal sessions was causing some physical discomfort, and by a review of audiological data, gathered by other medical services. Accordingly, the rehearsal facilities used by three of the bands, located in Calgary (1), Kingston and Halifax were reviewed.

METHODS

The approach to studying the facilities included three main aspects: the sound exposure experienced by the musicians, the acoustic parameters of the rehearsal studios, and the impressions of the musicians and directors as users of the facilities.

The levels of sound to which musicians are exposed during rehearsal vary with the position in the ensemble, the instrument played, the nature of the music, the technique of the director and the acoustics of the studio. Hence, estimating the sound exposure experienced by musicians requires measurements at a number of locations within the room, using an integrating dosimeter or sound level meter, to account for the time-varying nature of the music.

Accordingly, three dosimeters (five in Halifax) were used in the studies, and the microphones were attached to the epaulets of representative musicians. The dosimeters used a 5 dB exchange rate in accordance with current CF and Federal Government Department of Labour regulations (2). Monitoring was carried out over two consecutive days and the directors were asked to conduct the customary three-hour rehearsal sessions in as typical a manner as possible.

Several acoustic properties of the main rehearsal rooms were assessed to study the relationship between the practice space and sound exposure. Reverberation times were measured as a function of frequency in each room, and the size, the methods of construction and the use of acoustic materials were noted. From these data, the effects of alternate acoustic treatments on reverberation times were estimated.

RESULTS AND DISCUSSION

The results of dosimetry measurements are given in Table 1.

<table>
<thead>
<tr>
<th>Dosimeter Exposure Data for CF Musicians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 1</td>
</tr>
<tr>
<td>TIME TO ACCUMULATE 100 PER CENT EXPOSURE</td>
</tr>
<tr>
<td>Trombone</td>
</tr>
<tr>
<td>Trombone</td>
</tr>
<tr>
<td>Keyboards</td>
</tr>
<tr>
<td>Euphonium</td>
</tr>
<tr>
<td>Tube</td>
</tr>
<tr>
<td>Trumpet</td>
</tr>
<tr>
<td>Band 2</td>
</tr>
<tr>
<td>Trumpet</td>
</tr>
<tr>
<td>Trumpet</td>
</tr>
<tr>
<td>Alto Sax</td>
</tr>
<tr>
<td>Alto Sax</td>
</tr>
<tr>
<td>Trombone</td>
</tr>
<tr>
<td>Traps</td>
</tr>
<tr>
<td>Bassoon</td>
</tr>
<tr>
<td>Band 3</td>
</tr>
<tr>
<td>Trumpet</td>
</tr>
<tr>
<td>Alto Sax *</td>
</tr>
<tr>
<td>Trombone</td>
</tr>
<tr>
<td>Alto Sax *</td>
</tr>
<tr>
<td>Traps</td>
</tr>
<tr>
<td>Euphonium</td>
</tr>
<tr>
<td>Fender bass</td>
</tr>
<tr>
<td>Clarinet</td>
</tr>
<tr>
<td>French Horn</td>
</tr>
<tr>
<td>Trumpet</td>
</tr>
<tr>
<td>Guitar</td>
</tr>
</tbody>
</table>

* = Musician monitored for two days

The sample times ranged from 20 to 200 minutes, the average being 96 minutes. The resulting readings were normalized for Table 1 by calculating the time that would be taken for the musicians to acquire a 100 per cent daily noise dose, i.e., equivalent to 90 dBA for an eight-hour exposure. No distinction was made between the first and second day of rehearsal, because in only one instance was it possible to monitor the same individual during two consecutive sessions and concurrently, provide a useful cross-section of data for different
Despite considerable inter- and intra-band variability, the trumpet, trombone and saxophone sections appear to be the areas where the greatest risk of overexposure exists. With minor variations, the layout of these sections in each room was the same; trumpets in the back row, trombones in the centre and saxophones closest to the podium.

In several instances, it was noted that a 100 per cent noise dose accumulated before the sessions were finished, indicating a sound overexposure during these particular rehearsals. Because the sample size was small, one must use caution in extrapolating these results to the long term. Indeed, the one dosimeter fitted to the same musician on both days, showed a 26 per cent variation in exposure from one day to the next.

A useful parameter in understanding the influence of the rehearsal space on sound exposure is the "room radius", calculated from the dimensions of the studio and sound absorption data. A source radiating sound into an enclosed space generates two co-existing sound fields: a) the direct field, the sound pressure of which varies inversely as the distance from the source, and b) the reverberant field comprising reflected sound, the pressure of which is (theoretically) independent of position. The room radius is the distance from the source at which the magnitude of these fields is equal. Sound-absorptive materials control only the magnitude of the reverberant field energy, and have no effect on direct-path sound propagation. Hence, the larger the room radius will be the room's contribution to sound exposure.

The room radii for these band rooms ranged from 2.5 to 4.1 metres at mid frequencies. By visualizing the arrangement of musicians within a rehearsal space, and considering each instrument as a source, it becomes clear that most of the musicians are within the room radius (i.e., in the direct field) of at least some of the instruments. Only those beyond this distance receive the greater portion of their sound exposure from the reverberant field.

The room radius was recalculated for each band room, assuming that more effective room absorption could be introduced by additional sound-absorptive materials. Since these facilities were already reasonably absorptive, it was found that the assumed changes would produce only a modest increase in the room radii (about 20 per cent), and hence reduce the overall sound exposure for very few musicians.

In the interests of minimizing exposure to sound, and optimizing the available rehearsal space for musical purposes, the bands have experimented (with limited success) with a number of factors under their control. Some use risers to provide vertical isolation, or practice in sound- and/or acoustically shielded areas. All have tried alternate floor arrangements for their musicians and/or sections, and break periods of varying lengths during rehearsals. The concept of wearing hearing-protective devices has proven unpopular, due to the substantial change caused in the perceived sound, and the loss of ensemble that follows. A more effective overall approach might be to "hold back" on fortissimos during practice sessions, or to modify the repertoires to include a wider selection of quiet music.

Many of the musicians questioned during these evaluations stated that the sound of the bands seemed "too loud" or "too full" for the given rehearsal space, particularly during strong bass passages. In two of the facilities, the reverberation times did tend to increase with decreasing frequency, suggesting that the rooms indeed were too responsive at low frequencies (see Table 2).

### Table 2

<table>
<thead>
<tr>
<th>Room</th>
<th>Octave-Band Centre Frequencies (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>63</td>
</tr>
<tr>
<td>1</td>
<td>.49</td>
</tr>
<tr>
<td>2</td>
<td>.75</td>
</tr>
<tr>
<td>3</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Patrick et al (3) found that musicians expressed greater satisfaction with rehearsal facilities when the reverberation time was quite short and essentially independent of frequency. For musical reasons, therefore, it would seem useful to alter the acoustics of rooms two and three, by introducing more low-frequency absorption. None of the rooms surveyed contained resonant absorbers designed specifically for low-frequency sound reduction.

### CONCLUSIONS

In practice, few options are available to reduce the sound exposure experienced by musicians, that are also acceptable on musical grounds. Sound exposure depends on two variables; the sound pressure level, and the duration of exposure. It follows that a reduction in either quantity would serve to lessen the exposure.

The differences in exposures experienced by the band members surveyed in these studies were more likely attributable to the styles of music and direction, or day-to-day variation than to band layout or studio acoustics.

The fullness of sound noted by the musicians during bass passages may provide a musical reason for increasing the low-frequency absorption characteristics of certain CF rehearsal studios.

### REFERENCES


1. INTRODUCTION

Petković Ivan, "14.OKTOBAR"
Grđinić Gojko "14.OKTOBAR"

Razvojoj-opitni centar - 37000 KRUŠEVAC
YUGOSLAVIA

Dans les exigences du cadre de gestion de l'Industrie "14.octobre" Kruševac concernant la qualité suprême de ses machines du point de vue de certitude et de construction, et de celles de sa durée du service après vente et de l'entretien de ces machines.

Aux équipes des experts on a posé le problème qui était que l'homme doit être protégé au cours du travail et qu'il doit être au maximum protégé des influences négatives.

Les ambitions de l'Industrie "14.octobre" qui étaient que ses machines aident le ouvriers sur les chantiers de tous les continents et dans toutes les conditions étaient devenues la réalité.

Les normes les plus rigoureuses de SAE, DIN, ISO ainsi que les critères de OSHA sont satisfaits avec; la pelle mécanique ULM-160 B et les tracteurs à chenilles TG-140 et TG-220.

Pour réaliser cette mission on a utilisé les principes de base des sciences exactes composés dans un ensemble avec beaucoup d'esprit.

Les cabines représentent une unité intégrale et elle sont séparée du châssis par les amortisseurs. Elles sont thermiquement et de bruit isolées. Celles-ci sont étanche et avec la surpression à l'intérieur ce qui interdit la pénétration de la poussière et de l'eau dans la cabine ainsi que le courant d'air.

Les cabines sont aérées avec l'air filtré de très haut niveau d'épuration. Pour assurer le chauffage l'on a installé le calori-fère de très haute capacité.

A la demande du client on peut livrer la cabine avec l'air conditionné, les vitres en couleur la radio etc.

Toutes les commandes les instruments de contrôle de la poussière et le siège confortable sont disposés dans les zones optimum de manipulation où l'on a rempli les normes ergonomiques du confort.

Les cabines satisfont les conditions de protection du chauffeur au cas du renversement dans l'accident, de la chute des pierres ou des objets sur le toit (écroulement) (ROPS et FOPS).

L'attention particulière est consacrée à la limitation de bruit et de vibrations dans les niveaux souhaitables.

Présentons quelques réflexions techniques matérialisées à travers la solution des problèmes de protection contre le bruit et les vibrations et en même temps elles ont contribué à ce que nos machines ULM-160 B, TG-140 et TG-220 soient parmi les meilleures dans le monde sur le plan des rapports

l'homme - la machine.

Selon les résistances des appuis on a fait la sélection de la rigidité pour éviter la zone de résonance quit est pour ces machines de 4 à 11 Hz.

Pour diminuer le temps de réverbération dans les points précisément déterminés on a installé le matériau adéquat (comme isolant passif contre le bruit) et avec ça on a diminué considérablement le bruit selon le principe du resonateur de Helmholtz.

Mesure du bruit (son)

Mesure et analyse du bruit sont faite selon une chaîne de mesure présentée dans le texte, et selon les recommandations OSHA,ISO et SAE normes, utilisant l'équipement de mesure suivant:

- microphone B et K 4165 diamètre 13 mm
- cable de mesure B et K A0028
- multiplicateur B et K Z603
- Bell et Howell DCM K 11 - 80 -10
- inscripteur magnétique Bell Howell NAGRA TI
- analysateur structurale dynamique HP 5423 A
- calculatrice de tableau HP 9825 B
- Plotter HP 7225 B

Etat de mesure pour l'analyse du bruit est programmé sur HP 5423 A et présenté sur la figure No1.

\[\text{Figure 1.}\]

Le programme d'essai et d'analyse du bruit est réalisé par le Centre de développement et des essais de l'Industrie "14.octobre" Kruševac à travers les activités suivantes:

- mesure de bruit au poste de l'opérateur
- mesure de bruit d'émission auto: de la machine
- mesure de bruit d'émission lors du passage de la machine.

Mesure de bruit au poste de l'opérateur a été faite:

- par le test statique à nombre de tours max et avec le nombre de tours du moteur de 2000 moins 1
- lors du transport de la machine avec les portes ouvertes et fermées et à toutes les
puissances de transmission.

Lors du cycle de travail par les portes ouvertes et fermées:

Nombreuses données présentées dans cet exposé ainsi que les photos annexées sont liées à la machine ULT-160 B, car du fait de limitation du surface pour écrire nous n'étions pas en mesure de présenter en tel nombre TG-140 et TG-220. Beaucoup de détails liés à la machine TG-140 et TG-220 qui peuvent être obtenus auprès de l'auteur du présent texte ne se distinguent pas considérablement et toutes les trois machines satisfont les exigences des critères OSHA.

D'après les résultats d'éssais obtenus à travers des mesures, figure No 2.

et après leur analyse on voit bien qu'on a atteint un haut niveau de notre machines ULT-160 B figure No 3.

Le bruit autour des machines concernées (bruit d'émission) dont tous les 8 points de mesure à une distance de 15 mètres est inférieur à 86 dBA et par conséquent ces machines conviennent pour les travaux dans les milieux urbains respectant toujours les normes et les dispositions légales.

Le bruit mesuré au moment du passage des machines (ULT-160B, TG-140 et TG-220) à distance de 15 mètres est en dessous de 86dBA sorte qu'on a évite complètement la possibilité que le bruit soit génant pour la population.

Les raisons d'application de nos machines sont de la nature économique et sociale. Sur le plan social on diminue le manquement à la santé des ouvriers et par conséquent on diminue l'absentéisme dû à la maladie et à l'invalidité.

Les références économiques sont faciles à enregistrer. En premier lieu, c'est laissé de retour des effets sociaux, en suite le fait que dans la cabine confortable on ne sent ni trop de chaleur (en été) ni froid (en hiver) l'ouvrier reste plus longtemps au travail avec une motivation plus élevée et on augmente la productivité jusqu'à 23%. Et comme il s'agit des machines de grandes puissances, de 117 kw et de plus, il est très facile de démontrer la justification économique de leur utilisation.

CONCLUSION

1. Les résultats obtenus prouvent que sur le plan de diminution de bruit nous avons atteint le niveau mondial, cependant ces résultats doivent être améliorés de plus en plus au cours du travail permanent de développement et des recherches.

2. La nécessité d'adaptation au marché exige un comportement souple dans toutes les conditions et par rapport à toutes les exigences qui se présentent en un jour en jour deviennent plus rigoureuses et plus complexes. Pour cette raison nous offrons à nos clients la possibilité de faire le choix des machines selon leurs souhaits.

3. Le confort, la sécurité, la certitude le service après vente très réparti et les prix attractifs sont les caractéristiques principales des machines du programme de fabrication de l'industrie "14 octobre" Kruševac.

4. En plus du sentiment personnel de l'ouvrier lors du travail avec nos machines (travail sans reproches aux conditions de travail et de santé) vous n'ourez aucune d'arière législative.

5. Le syndicat ne pourra formuler aucune reproche concernant les conditions de travail et la puissance du moteur sur ULT-160 B de 117 kw et propres fabrications des bottes à vitesse, des ponts et des autres pièces vitales assurant un fonctionnement et travail sûr et sans gène pendant plusières années aux conditions de travail les plus difficiles.

6. Il suffit que votre opérateur sente dans la cabine de notre pelle mécanique ULT-160 B pour avoir le sentiment que tout est agréable et facile pour manipuler avec un rendement très élevé.
A UNIVERSAL APPROACH TO ALIGNMENT OF NOISE EXPOSURE MEASUREMENTS AND DATA

Terrence A. Dear
E. I. du Pont de Nemours and Co., Inc.
Wilmington, Delaware 19899

INTRODUCTION

Noise criteria for the workplace usually are set for one major purpose -- to protect the hearing health of individuals, specifically regarding preservation of their ability to communicate human speech. Thus, it is alarming to see the great differences in criteria, dose and exposure time relationships that exist among various standards around the world. It must be recognized that implementation and compliance costs also vary widely according to criteria specified in each rule and corresponding enforcement procedures.

Exposure evaluation is critical to demonstration and documentation of regulatory compliance. Instrumentation requirements and sound level measurements must be in complete harmony with numerical noise criteria limits and their historical development. This is essential in order that comparisons of measurements with criteria yield valid representations of exposure risk.

During the 1970's exposure evaluation evolved from simple, almost identical instruments to include a new wave of instrumentation introducing variations well beyond the stringent specifications of governing standards. Regulatory pressures to go from "single number" workplace measurements made with the traditional instruments, upon which Damage Risk Criteria (DRC) were based, to complex integrations of dose, introduced many new concerns and needs. Tools and procedures outlined in this monograph address one of the major needs in context of corresponding regulations.

General formulas are presented, with examples, that can facilitate data translation from one set of instrumentation and criteria to another. These tools have been used effectively in comparisons of data from European plants with Company standards and U. S. Government (OSHA) requirements. Accordingly, we have reduced capital investment in new instruments and avoided costs of redoing noise surveys.

UNIVERSAL APPROACH: BACKGROUND TO DEVELOPMENT OF FORMULAE

In early 1970, the author developed a simple mathematical relationship (an original dosimeter algorithm) describing OSHA's regulatory criteria for noise exposure:

\[ T = \frac{S}{L - L_0} \]  

(1)

T and S are in same units of time.

L is the exposure duration, L_0 is the exposure sound level, L_0 is the criterion level and the exchange rate. T is permissible exposure time.

The formula was given to NIOSH in 1971, and OSHA's version of this formula was published in the 1974 proposed regulation. The universal formulae that follow were part of this original development.

UNIVERSAL APPROACH: EQUIPMENT STANDARDS

This approach assumes that each instrument explicitly meets ANSI S1.4^2 specifications (or equal).

UNIVERSAL APPROACH: DERIVATION OF EQUATIONS

The original published equations were based on:

\[ L = L_0 + \frac{10R}{3} \log \frac{D}{125 S} \text{ dBA ref 20 } \mu \text{Pa} \]  

(2)

where D is dose exposure and S is in hours.

Note that in Equation (2) the argument of the logarithm permits statistically valid samples of dose D to be used. In that case, S is the appropriate sampling time during which D was acquired.

For two qualifying instruments designated by subscripts 1 and 2, Equation (2) yields two equations for S = 8 hrs (S can be variable, in general).

\[ L_1 = L_{c1} + \frac{10R_1}{3} \log \frac{D_1}{100} \]  

(3)

\[ L_2 = L_{c2} + \frac{10R_2}{3} \log \frac{D_2}{100} \]  

(4)

Setting \( L_1 = L_2 \) yields:

\[ L_{c1} - L_{c2} = \frac{10R_2}{3} \log \frac{D_2}{100} - \frac{10R_1}{3} \log \frac{D_1}{100} \]  

(5)

Further rearrangement gives:

\[ D_1 = (D_2)^{R_2/R_1} \frac{10^{R_1}}{10^{R_2}} \]  

(6)

For example, one application involves relating doses measured with an instrument for which \( L_{c2} = 90 \) dBA and \( R_2 = 3 \) dBA to dose measurements in a system where criteria are \( L_{c1} = 90 \) dBA and \( R_1 = 5 \) dBA.

The resulting compact expression is:

\[ D_1 = 10^{45} (D_2)^{3/5} = 6.31 (D_2)^{3/5} = 6.31 (D_2)^{0.6} \]  

(7)

Also useful is an equation relating permissible time of exposure, T, and dose, D, for any set of \( L_c \) and R (S is in hours).

\[ T = \frac{S}{10} \]  

(8)

The following is one example of how to use these relationships.

1. Given a noise dose measurement \( (D_2) \) made with a qualified instrument based upon one set of criterion level, \( L_{c2} \) and exchange rate \( R_2 \). It is desired to calculate the noise dose \( (D_1) \) and permissible time of exposure \( (T_1) \). That would pertain to compliance with a regulation based upon another criterion level and exchange rate set \( (L_{c1}, \text{ and } R_1) \).

2. Use Equation (6) to obtain \( D_1 \) in terms of \( D_2 \) and compute \( T_1 \).

3. Use Equation (8) to compute the permissible time of exposure \( T_1 \) for \( D_1 \) and S.

Convenient tables and charts have been made relating various sets of criterion levels and exchange rates using these expressions. These procedures are valid on the basis that the integrated time/level output of any qualified exposure measuring instrument gives a total value for which the representative sound level, sometimes referred to as the "time weighted average (TWA)," can always be calculated [eq. by Equation (2)].

Figure shows the relationship between doses and sound exposure levels for the two sets of criteria of Equation (7).
For example, if $D_2$ is 1000 units (i.e., $L = 100$ dBA), then $D_1$ is about 400 at 100 dBA. Incredibly, a 2.5 to 1 dose ratio exists for the same level of exposure between two different measurement systems assuming the same stated risk control goals. The difference in permissible exposure times for the two systems in this case is 1.2 hours, using Equation (8). Note that the ratio of permissible exposure times is also 2.5 to 1.

Many more comparisons can be generated using the relationships in this excerpt of a detailed monograph on the subject.

CONCLUSIONS

Requirements for measurement instrumentation and application thereof are typically the most significant technical specifications in noise regulation. As the primary determinants of compliance status, exposure measurements are key factors in assessing the total cost impact of noise regulation. It has been shown that differences between instruments using various criteria sets are extremely large and significant. Practically, problems may arise, including liabilities, for specifiers and users of instrumentation that is not in complete harmony with historical data bases for Damage Risk Criteria (DRC)\(^3\) and related noise limits.

Equations have been presented that can facilitate conversions and comparisons between various noise criteria and the data gathered with qualified\(^2\) instrumentation.

Regarding differences caused by selection of exchange rate ($R_i$). W. D. Ward noted\(^4\) that the equal energy principle and its associated 3 dBA rate do not fit the patterns of intermittent exposure to noise typical in North American Industry. Sulkowski\(^5\) reached a similar conclusion based upon a comprehensive study of existing data.

The magnitude and impact of the differences among various sets of criteria, for example, regarding permissible times of exposure, $T$, at respective sound levels, seem far out of alignment with typical historical comparisons of supporting data bases for each criteria set\(^6\).

For example, substituting an exchange rate $R$ of 3 dBA for an existing $R$ of 5 dBA in an enforced standard can impose serious economic penalties that are inappropriate and unnecessary to achieve the common hearing conservation objectives stated at the outset. A 90 dBA criterion level for 8 hours, with an exchange rate of 5 dBA, is demonstrably the most practical and effective long-term approach to setting regulatory criteria for workplace noise\(^8\). New epidemiological evidence confirms this\(^1\).

ACKNOWLEDGEMENT

This paper is dedicated to my friend and colleague Alfred Lechner who, to the best of my knowledge, had the first working model of what is today called a noise dosimeter in the late 1960's and was duly awarded a patent for the device. In addition, he coined the term "noise dosimeter" in that era.

EFFECT OF THE NUMBER OF MEASUREMENTS
OF SOUND INTENSITY ON THE ACCURACY
OF SOUND POWER ESTIMATION

Mei Q. Wu and Malcolm J. Crocker
Department of Mechanical Engineering
Auburn University, AL 36849, U.S.A.

In sound power determination via sound intensity measurements, two sampling techniques are
usually used to approximate Gauss' integral: 1) discrete stationary sampling at fixed points on the
measurement surface, and 2) continuous scanning by moving the probe over the surface. Comparisons
between the two sampling methods have been studied by Buffa and Crocker [1] and Bockhoff [2]. In this
paper the properties of the estimation error introduced by approximating the integral by a
finite number of samples is studied. In particular, the dependence of the error on the number of samples, the distance between the source and the measurement surface, and the location of the
source with respect to the sampling points was examined.

THEORETICAL FORMULATION

Let us consider a sound source element $Q$ at
point $(u, v)$ on a plane $B$, and a receiving point $P$
at $(x, y)$ on a parallel plane $A$. The normal
distance between the planes is $L$. (Fig. 1)
The velocity potential at point $P$ caused by
element $Q$ is

$$
\phi(u, v, x, y) = \frac{1}{4\pi r} e^{i(\omega t + \theta - kr)}
$$

where $\phi$, $\omega$, $k$ and $r$ are all functions of $u, v, x, y$.

At point $P$, the total potential caused by
all the elements distributed on plane $B$ is

$$
\phi(x, y) = \int_B \frac{1}{4\pi r} e^{i(\omega t + \theta - kr)}
$$

dudv.

The sound pressure at $P$ is

$$
p(x, y) = \rho \frac{d\phi}{dt} = \frac{\omega}{4\pi} \int_B \frac{1}{r} e^{i(\omega t + \theta - kr)}
$$
dudv

and the particle velocity at $P$ in the $z$ direction is

$$
u(x, y) = -\frac{\omega}{4\pi} \int_B \frac{1}{r} e^{i(\omega t + \theta - kr)}
$$
dudv.

The time averaged irrotational sound intensity
in the $z$ direction is

$$
I(x, y) = \frac{1}{2} \text{Re}(p^* u^*)
$$

where $I(x, y) = \frac{1}{2} \text{Re}(p^* u^*)$

$$
\int_B \frac{1}{r} \left\{ \cos(\theta - kr) - k \sin(\theta - kr) \right\} dudv
$$

$$
+ \int_B \frac{1}{r} \left\{ \cos(\theta + kr) - k \sin(\theta + kr) \right\} dudv
$$

$$
\int_B \frac{1}{r} \left\{ \sin(\theta + kr) \cos(\theta + kr) \right\} dudv.
$$

Eq. (7) gives a general expression for the
normal sound intensity caused by source elements
which may have different magnitudes and initial
phases. Using Eq. (7), the exact sound power
passing through the plane area $A$ and the
approximate power obtained from a finite number of
measurements of the sound intensity on the plane $A$
can be calculated.

NUMERICAL SIMULATION AND CONCLUSIONS

When there is only one monopole on plane $B$, Eq.
(7) can be simplified to

$$
I(x, y) = \frac{1}{2\pi} \int_0^\infty \frac{\omega}{4\pi r^2} \left\{ (x-u)^2 + (y-v)^2 + L^2 \right\}^{-3/2}
$$

dudv.

Let the measurement surface $A$ be a square of
side length $L$. The exact power $W$ passing through
the measurement surface $A$ is

$$
W = \frac{\omega}{2\pi} \int_0^\infty \frac{\omega}{4\pi r^2} \int_0^\infty \left\{ (x-u)^2 + (y-v)^2 + L^2 \right\}^{-3/2}
$$

dudv.

The approximate power $W$ estimated by a finite
number of sound intensity measurements is given by:

$$
W = \sum_{i=1}^N I(x_i, y_i) S_i
$$

where $I(x, y)$ is the normal sound intensity at
sampling point $(x_i, y_i)$, $S_i$ is the area element, and
$N$ is the number of sampling points. Let the
approximate sound power $W$ be estimated successively
by $L = 2, 4, 6, 9$ and $16$ measurements points in a $1 \times 1$

square. The distributions of the sampling points are shown in Fig. 2.

The estimation error is defined as

$$
E = 10 \log \left( \frac{W}{W} \right).
$$

Calculations of the error were made for various
source positions, distances $L$ and numbers of
sampling points. Examining the results plotted in
Figs. 3 through 8 gives the following conclusions. 1.
When the number of sampling points is very
large, say 16 points in a $1 \times 1$ square, the
estimation errors for most cases are less than 1
db, no matter how close the measurement surface is
to the source.
2. When the distance between the measurement
surface and the source is twice as large as the
length of the side of the measurement surface, the error is always less than 1 dB, no matter how many measurements are made; and when L tends to infinity the error tends to zero.

3. When 0.25 L/d > 1.5, the use of more sampling points reduces the error. Thus the accuracy of the sound power estimation can be improved by increasing the number of intensity measurements as long as the measurement surface is not too close to the source.

4. When the measurement surface is very close to the source, say L/d < 0.25, the error depends on the location of the source with respect to the sampling points. It is not necessarily correct to assume that the use of more sampling points will make the sound power estimation more accurate. The accuracy of the power estimation cannot be improved by increasing the number of sound intensity measurements, unless the number is very large, say 16 points in a 1m x 1m square.

EXPERIMENTAL TESTS

Some experiments were carried out to check the theoretical curves shown in Fig. 3a. The experimental results are shown in Fig. 4. It can be seen that the experimental results agreed well with the curves calculated from the simulation equation except at a few points.

REFERENCES


Fig. 1. The source element and the receiving point.

Fig. 2. The distributions of the sampling points.

Fig. 4: Comparison of the experimental results with the curves in Fig. 3a. The solid lines are the curves in Fig. 3a, and o, x and □ denote the experimental results obtained by 1, 4 or 9 points, respectively.
SELECTIVE CONDITIONING OF ACOUSTIC INTENSITY MEASUREMENTS FOR SIMPLE AND MULTIPLE SOURCES

P.R. Wagstaff & R. Pech

Université de Compiègne, B.P. 233, 60206 Compiègne, France
**E.N.S.T. rue Barrault, 75013 Paris, France

INTRODUCTION

Acoustic intensity measurements may be used to determine the sound power of industrial noise sources in situ providing the effects of nearby sources, problems of access and the reflection effects are relatively limited. In cases where several sources are grouped together, and where the measurement surface does not completely enclose the source, the accuracy of the result is prejudiced. The use of intensity measurements to determine the contribution of each source in a closely grouped multiple source situation is therefore unsatisfactory and the use of selective conditioning techniques would seem to provide a solution to this problem. The contribution of each source could then be determined with a surface enclosing several sources and in cases where the intensity measurements are used to assess the sound power radiated by different parts of the same machine the results would be less perturbed by the interactions between the parts. The use of two or three dimension measurements of intensity to provide ancillary information to locate the sources would also be more reliable information in the presence of disturbing sources if the conditioning was applied to select a single source and suppress reflections.

CONDITIONING TECHNIQUES

The cross spectral formulation to calculate the acoustic intensity [1] uses the direct cross spectral density between two closely spaced microphone signals \( G_{12} \), and the density of the medium \( \epsilon \).

\[
I = \frac{\text{Im}( G_{12} )}{\epsilon \times w \times d} \tag{1}
\]

where \( \text{Im} \) denotes the imaginary part, \( w \) the angular frequency and \( d \) the distance between the microphones.

The use of a source reference signal permits a calculation of the cross spectrum due uniquely to the part of the signal that is coherent with the source reference signal giving a selective intensity.

\[
I_{s} = \frac{\text{Im}( H_{1}^{*} \times H_{s} \times G_{ss} )}{\epsilon \times w \times d} \tag{2}
\]

where \( H_{1} \) and \( H_{s} \) are the transfer functions between the source reference and the first and second microphone signals respectively, * denotes the complex conjugate and \( G_{ss} \) is the autospectral density of the reference signal.

For investigations of uncorrelated sources the selective intensity may be calculated by the channel ‘simultaneous’ processing or ‘sequential’ dual channel processing of the source signal and each microphone in turn. In the latter case the second microphone may be dispensed with and the second transfer function may be obtained by displacing the first microphone. This avoids matching problems and diffraction effects that introduce errors at low and high frequencies respectively with the two microphone techniques. Time windows may also be applied to condition out the reflections due to the surfaces surrounding the source, provided that the source is not deterministic and that the time between the arrival of the direct wave and the first reflection is sufficient to permit a window giving reasonable resolution. The window is then chosen to be shorter than this time, the source reference signal is delayed to correct for the propagation time of the direct wave and the reflections are completely uncorrelated with the reference signal and can be eliminated by sufficient averaging. The technique requires the use of relatively short record lengths which limits the resolution of the results and the no longer stationary of the signals is reduced by the rejection of the reflected waves which implies that a greater number of averages is required to converge to an accurate result.

The processing techniques that are required for correlated sources differ mainly in the calculation of the transfer functions, where residual spectra estimations are used, and in the requirement of simultaneous acquisition of all the input and output signals to reduce the risk of error [2]. The source reference autospectra may be used directly with those transfer functions to calculate the intensity if it is certain that there is no inter-source contamination of the reference transducers. In cases where the coherence is due entirely to the effects of contamination the use of the residual autospectrum for each source will at least give a lower limit estimation for the intensity due to each source.

APPLICATIONS AND RESULTS

Selective techniques of intensity measurement are in an early stage of development and limited to laboratory test and simulations. Investigations of coherent sources are more subject to error than independent source measurements and the results presented here are limited to the latter case. Transmission loss measurements have been carried out in conditions of high background noise (due to a second source) with selective intensity measurements. An increase of the efficiency of the noise rejection is given in the figures 1 to 3 for an intensity measurement using the noise generator signal as a reference [3].

Results for the suppression of reflections are shown in figures 4 to 10 after an investigation of the cross correlation and transfer function obtained for a normal and a reduced time window. The reflections shown in the cross correlation in figure 4 are eliminated in figure 6 by the time window and the resulting transfer function is given in figure 8. The source in this case was a loudspeaker 1.3 m from the floor and the measurements were made on the axis at a distance of 1.5 m. Figure 10 shows the reflection free intensity measurement which is better than the anechoic room result which was itself modified by an echo from the track of a moveable platform.

In conclusion it may be said that these methods can be made to work and that the increase in complexity of the technique is justified by the increased precision of the results that it can provide.

REFERENCES

Figure 1 Direct intensity with source one in action.

Figure 2 Direct intensity with both sources in action.

Figure 3 Selective intensity of source one, both sources in action.

Figure 4 Cross correlation between source and first microphone position.

Figure 5 Modulus of transfer function.

Figure 6 Cross correlation with time window applied.

Figure 7 Modulus of transfer function with time window applied.

Figure 8 Intensity measured in reverberant room.

Figure 9 Intensity measured in anechoic room.

Figure 10 Intensity measured in reverberant room with time windows applied.
ANTENNE INTENSIMETRIQUE POUR LA MESURE DES CARACTÉRISTIQUES DE RAYONNEMENT DES STRUCTURES VIBRANTES

P.J.T. Filippi, D. Habault, et J. Pireaux
C.N.R.S. Laboratoire de Mécanique et d’Acoustique
31 chemin Joseph Aiguier, B.P. 71
13277 Marseille Cedex 9, France

I-INTRODUCTION

Considérons une source sonore occupant un domaine Ω de frontière Ω, et émettant un signal harmonique \( \exp(\text{i} \omega t) \). La pression acoustique \( P(M) \), rayonnée en champ libre, peut être décrite de diverses façons, en particulier : a) par une série d’harmoniques sphériques centrés en un point arbitraire \( O \); b) par une intégrale de Green en fonction des valeurs de la pression et de la vitesse partielle normale sur la surface de la source (ces deux fonctions ne sont pas indépendantes). Ces deux représentations du champ sont absolument générales, et s’expriment à partir de quantités à peu près directement accessibles à l’expérience et d’une interprétation physique simple.

Au § II, nous définissons des problèmes de minimisations attachés à chacune des deux représentations du rayonnement d’une source, et en nous déduisons le protocole de mesure qui permet d’accéder aux paramètres intervenant dans ces modèles. Au § III, nous décrivons les résultats expérimentaux obtenus en chambre anéchoïque : les paramètres identifiés sont les coefficients de la série d’harmoniques sphériques. Au § IV on donne un exemple simple d’identification des paramètres intervenant dans la représentation de Green.

II-PROBLÈMES DE MINIMISATION, ET PROTOCOLE DE MESSURE

Soit \( O \) l’origine d’un système de coordonnées sphériques, et \( M(R,r,\theta) \) un point quelconque extérieur à \( \Omega \), la plus petite sphère centrée en \( O \) et contenant \( \Omega \). On sait que la pression acoustique créée par la source s’exprime par la série suivante :

\[
(1) \quad P(M; a, b) = \sum_{n=0}^{\infty} M_n \cos b_n \sum_{m=0}^{n} m_n \sin b_n \sin b_n
\]

\[
\times \left( \left( \sum_{m=0}^{n} M_n \cos b_n \sum_{m=0}^{n} m_n \sin b_n \sin b_n \right) \sin b_n \right)
\]

Dans cette expression, \( h(\mathbf{r}) \) est la fonction de Hankel sphérique d’ordre \( n \), et de première espèce ; \( P(z) \) est la fonction de rang \( n \) et d’ordre \( n \) de Legendre.

Soit \( C = \exp(\text{i} k R) / 4\pi \) la fonction de Green de la sphère de l’équivalent Helmholtz en espace indéfini. Si \( P \) désigne la pression acoustique sur \( \Omega \), et \( v \) la composante normale de la vitesse partielle, la pression rayonnée par la source est écrite :

\[
(2) \quad P(M; v) = \sum_{n=0}^{\infty} \frac{\sin b_n}{n!} \left( \left( \sum_{m=0}^{n} M_n \cos b_n \sum_{m=0}^{n} m_n \sin b_n \sin b_n \right) \sin b_n \right)
\]

\[
\times \left( \left( \sum_{m=0}^{n} M_n \cos b_n \sum_{m=0}^{n} m_n \sin b_n \sin b_n \right) \sin b_n \right)
\]

où \( v \) est le vecteur unitaire normal à \( \Omega \), et intérieur à \( \Omega \), et \( p \) est la masse volumique du fluide. On sait que \( p \) et \( v \) ne sont pas indépendants : pour fixer les idées on peut choisir \( v \) comme paramètre ; \( \exp(v) \) est alors considéré comme une fonction de la vitesse normale de l’enveloppe externe de la source.

Supposons qu’on connaisse \( P(M) \) sur une surface \( \Sigma_1 \) entourant la source et contenant \( \Sigma_1 \); choisissons pour simplifier une sphère de rayon \( R \). On peut déterminer les résultats suivants :

Il existe une et une seule suite \((a_{nm}, b_{nm})\) qui minimise la fonctionnelle quadratique suivante :

\[
(3) \quad J_1 = \sum_{n,m} \left( P(M) - P(M(a_{nm}, b_{nm})) \right)^2
\]

Il existe une et une seule fonction \( v \) qui minimise la fonctionnelle quadratique suivante :

\[
(4) \quad J_2 = \sum_{n,m} \left( P(M) - P(M(a_{nm}, b_{nm})) \right)^2
\]

Du point de vue expérimental, on n’a accès qu’à des données discrètes \( P(M_i) \), \( i = 1, \ldots, N \), de la pression rayonnée par la source. On est alors conduit à remplacer les fonctionnelles \((3)\) et \((4)\) par des expressions discrétisées ; de plus, on ne pourra déterminer qu’un nombre \( N \times N \) de paramètres pour décrire la source : la représentation \((1)\) est remplacée par une série tronquée ; dans \((2)\), \( v \) est approché par des éléments fins de frontière.

Le nombre minimum de points de mesures, et leur espacement sont déterminés par le théorème de Shannon. En effet, sur la sphère de rayon \( R \), la série \((1)\) représentant la pression sonore est une série de Fourier double par rapport aux paramètres \( n \) et \( m \). Sur le plan expérimental, il est nécessaire de procéder tout d’abord à un examen qualitatif du diagramme de directivité de la source : on peut ainsi mettre en relief la fréquence spatiale la plus élevée qu’il est nécessaire de prendre en compte ; la fréquence d’échauffement du champ est double de la précédente.

III-IDENTIFICATION DE LA REPRESENTATION MULTIPOLaire D’UNE SOURCE

Nous avons réalisé la source expérimentale suivante : deux chambres de compression, assemblées dos à dos, et munis de tubes de faible diamètre produisent chacune d’elles un rayonnement monopolaire ; une paire de chambres de compression munies de tubes de faible diamètre et accolées permet d’obtenir un rayonnement dipolaire. Une telle source offre la possibilité de réaliser des directivités très variées.

Nous avons réalisé une antenne de 6 microphones, répartis sur un arc de cercle de rayon de 1.69 cm. La source est placée sur une table tournante, le plan des trois sources élémentaires étant perpendiculaire au plan de l’antenne. En faisant tourner la source avec un pas angulaire de \( \pi / 11 \), on effectue des mesures de pression acoustique en 125 points répartis sur une demi-sphère. Ces mesures (amplitude et phase) sont rapportées à celle de la pression reçue par un microphone solidaire de la source.

A partir de ces mesures, nous avons fait deux types d’identification : on approche le champ expérimental par celui d’une source multipolaire, centrée au centre de l’antenne, et contenant les harmoniques jusqu’à l’ordre \( 5 \) ; on cherche la position optimale de 3 sources multipolaires d’ordre \( 2 \), dont on calcule les composantes.

Sure la fig.1 nous avons tracé le niveau de pression émis dans le plan des 3 sources élémentaires \( (-) \). Nous avons porté les points de mesures ayant servi à faire l’identification \( (-) \), le champ reconstitué à partir de la première identification \( (+) \), et celui correspondant à la seconde \( (\times) \). L’accord entre champ expérimental et champ modélisé est excellent : le modèle à 3 sources étant toutefois un peu meilleur. La fig.2 montre le champ normalisé calculé à partir de la représentation multipolaire de la source.

IV-IDENTIFICATION DE LA REPRESENTATION DE GREEN D’UNE DU CHAMP RAYONNÉ PAR UNE PLAQUE VIBRANTE

On envisage l’exemple bidimensionnel suivant : une bande élastique de largeur \( 2a \), encastrée dans un baffle indéfini parfaitement rigide, animée d’une vitesse normale \( v \) donnée. Rayonne une pression acoustique dans un fluide occupant un demi-espace. On calcule cette pression sur un réseau de \( N \) points partant sur une demi-ellipse de grand axe supérieur à \( 2a \). A partir de ces données, on cherche à approcher la vitesse \( v \) par une
fonction polygonale à QxN paramètres, en minimisant la fonctionnelle $J_3$ discrétisée : la résolution du système surdéterminé correspondant se fait par une méthode de Housholder. Les résultats présentés sur la fig.3 correspondent à des données réparties sur différentes courbes : cercle de rayon $2a$ (*); ellipse d’axes $4a$ et $a$ (•); ellipse d’axes $3.2a$ et $2.9a$. On voit que le résultat le meilleur est obtenu pour l’antenne la plus proche de la structure.

V-CONCLUSION

La méthode que nous venons de décrire permet d’accéder, à partir de mesures de pression acoustique, à toutes les caractéristiques utiles d’une source monochromatique : de la représentation (1), il est aisé de déduire la puissance totale rayonnée par la source, son diagramme de directivité à l’infini, et l’évolution spatiale du vecteur intensité. La représentation (2) permet, en outre, de déterminer les zones de la structure qui débitent le plus d’énergie; ce type de renseignements est fondamental pour optimiser le traitement acoustique d’une source de bruit.

Il n’est pas toujours possible de faire des mesures en chambre anéchoïque. On peut envisager deux méthodes pour s’affranchir des réflexions parasites. Dans le cas de la réflexion sur un plan absorbant, on introduit le modèle de sol dans la représentation du champ total, ce qui modifie en conséquence les représentations (1) et (2). Dans le cas où une modélisation simple des réflexions n’est pas possible, on effectue des mesures du champ sur 2 surfaces concentriques : il est alors possible de séparer champ direct et champ réfléchi.

Dans le cas d’une source émettant un bruit à large bande, on peut faire appel à l’analyse de Fourier pour se ramener au régime monochromatique. On peut également envisager un traitement direct : la série (1) par exemple est remplacée par une série d’harmoniques sphériques pour l’équation des ondes : les coefficients $\alpha$ et $\beta$ sont remplacés par des opérateurs de convolution que l’on peut déterminer par une méthode de filtrage.

BIBLIOGRAPHIE


**figure 1**

**figure 2**

**figure 3**
RAYONNEMENT ACOUSTIQUE DES STRUCTURES MÉCANIQUES

F. Guisnel : E.D.F./Département acoustique
P. Smadja 1, avenue du Général de Gaulle
92141 Clamart - FRANCE -

1. Introduction.

Le rayonnement acoustique initié par les vibrations des structures mécaniques contribue, pour une grande part, au niveau sonore élevé dans les salles des machines des centrales nucléaires. Malgré la complexité du phénomène, de nombreuses études tendent à montrer que les corps de turbine basse pression des turbo-alternateurs sont les principales sources de bruit à 100 Hz. Afin de mieux simuler le processus de rayonnement et de juger de l'efficacité de dispositifs insominants, nous nous intéressons à la modélisation vibratoire et acoustique des corps de ces turbines. Cet article présente un code de rayonnement acoustique des structures vibration et son application au calcul du champ rayonné par une turbine à vapeur basse pression.

2. Technique d'optimisation.

La technique proposée, basée sur des modèles de vibrations mécaniques (code d'éléments finis de structures ou programme d'analyse modale vibroacoustique) et de rayonnement acoustique, doit optimiser de façon itérative la conception de structures mécaniques du point de vue de leur rayonnement sonore (figure 1). Le modèle de rayonnement acoustique est au centre de cette approche : il est conçu pour traiter aussi bien des données théoriques (issues de modèles d'éléments finis) que des résultats expérimentaux (issues d'une analyse modale vibroacoustique).

![Figure 1: Optimisation du bruit rayonné par une structure. Optimization of the sound radiated by a structure.](image)

3. Modèle de rayonnement acoustique : "VARIA".

La relation entre les vibrations de structures et le bruit rayonné est complexe. Elle est donnée par la solution de l'équation de Helmholtz avec certaines conditions aux limites. Il apparaît que seules des méthodes numériques telles que les équations intégrales permettent de résoudre le problème dans un domaine infini et pour des géométries rayonnantes de forme complexe. Cependant ces méthodes présentent deux inconvénients : des difficultés numériques liées au calcul d'intégrales singulières et l'obtention d'un système final non symétrique (Sayh et Dusset, Schenck II/...). C'est pourquoi nous nous sommes orientés vers une formulation variationnelle par équations intégrales qui élimine les défauts précédents (Hamdi II/2).

Présentation de la méthode :

Soit à résoudre l'équation de Helmholtz :

\[ \Delta \phi + k^2 \phi = 0 \text{ dans } \Omega = \mathbb{R}^3 \setminus S_1 \cup S_2 \]

avec les conditions frontières suivantes :

- pressions imposées sur S1
  \[ \phi \big|_{S_1} = f \]
- vitesses imposées sur S2
  \[ \frac{\partial \phi}{\partial n} \big|_{S_2} = g \]
- condition de Sommerfeld à l'infini
  \[ \lim_{r \to +\infty} r \left( \frac{\partial \phi}{\partial r} - ik \phi \right) = 0 \]

avec \( f, g \) : fonctions connues
\[ \frac{\partial \phi}{\partial n} \] : dérivée normale
\( k \) : nombre d'onde

![Diagram](image)

Les inconnues auxiliaires \( \sigma, \mu \), définies par les conditions aux frontières S1 et S2 minimisent la fonctionnelle :

\[ L(\sigma, \mu) = \frac{1}{2} \left( \sigma - B \sigma + C \left( \sigma + 1 \right) D \sigma + 1 \right) + c \left( \mu - U \right) \]

où \( B, C, D, U, V \) sont des matrices qui ne dépendent que des géométries de S1, S2 discrétisées par éléments fins de surface, et des fonctions f et g.

En minimisant \( L \), nous aboutissons à un système linéaire symétrique à résoudre :

\[ \begin{pmatrix} B & C \end{pmatrix} \begin{pmatrix} \sigma \\ \mu \end{pmatrix} = \begin{pmatrix} U \\ V \end{pmatrix} \]
4. Calcul du rayonnement d'un corps de turbine basse pression.

Le code "VARIA" a permis le calcul du champ acoustique rayonné à 100 Hz par un corps de turbine basse pression. "VARIA" demande de connaître la géométrie de la structure (surface S2) ainsi que son champ vibratoire de surface (fonction g) et fournit alors les niveaux sonores rayonnés en tout point de l'espace extérieur.

Le calcul est effectué en champ libre, les niveaux vibratoires ont été mesurés en 374 points du corps de turbine (valeurs en module et phase). Le corps est modélisé à l'aide de 544 éléments triangulaires plans (figure 2). La carte calculée des niveaux de pression rayonnés est donnée figure 3.

![Figure 2](image)

Description d'une enveloppe de turbine basse pression - Maillage utilisé.
Description of the outer cover steam turbine mesh used.

![Figure 3](image)

Carte de pression calculée. Rayonnement d'un corps de turbine.
Calculated sound level map. Radiation of an outer cover steam turbine.

5. Conclusions.

Le code de rayonnement "VARIA", basé sur une formulation variationnelle par équations intégrales, permet de déterminer avec précision le champ sonore rayonné par une structure. Couplé à des codes de vibrations de structures, il constitue un outil intéressant de prédiction de niveaux sonores rayonnés, de simulation acoustique de modifications structurelles et de systèmes insonorisants.

Bibliographie.

/2/ M.A. HAMDI. "Formulation variationnelle par équations intégrales pour le calcul de champs proches et lointains", Thèse de doctorat d'État (1982), UTC - COMPIEGNE - FRANCE.
MULTIPLE INPUT MODELS APPLIED TO INDUSTRIAL NOISE SOURCES - EXPERIMENTAL EVALUATION

D. Taillifet *, P.R. Wagstaff **, J.C. Henrio **
* Département Acoustique, E.D.F. - DER, 1 avenue du Général de Gaulle 92141 Clamart, France
** Université de Compiegne, B.P. 233, 60206 Compiegne, France

INTRODUCTION

Industrial noise sources may be investigated using a variety of measurement techniques. The major difficulty in most cases is that satisfactory values may be obtained for the noise spectra and the sound power of the sources, but the interaction with the surrounding environment, particularly in closed spaces such as industrial buildings is difficult to quantify. An excessively high noise level in a given area which is surrounded by several sources of more or less the same sound power may be due to only one of them if it is placed in a position that directs most of the sound power towards this area. The use of sound power measurements is therefore insufficient in such cases and must be linked to a calculation of the effects of the acoustics of the building. If direct measurement techniques are required, the use of shielding or coherence techniques are the only valid solutions of the problem of source identification, in the case of differences in the spectra of the sources. The use of coherence techniques is apparently much simpler than shielding techniques because they require representative signals of the sources rather than the complete suppression of their effects. This paper describes the use of such methods to locate the most important sources in an area of high noise levels in a nuclear power plant and in a reverberant room.

THEORETICAL PRINCIPLES

In a multiple source situation, the characteristics of the sound radiation of the sources may be represented by coherence functions obtained using microphones in the near field of each source or accelerometers placed on the vibrating surfaces. The effect of each of these sources at the output measurement point may then be expressed in terms of transfer functions between the input reference signals and the output. The complete set of cross and auto-spectral densities of the inputs and the output has to be calculated. The transfer functions may then be obtained using the following relationship [1]. For n sources:

\[ G_{ij} = \sum_{k=1}^{n} G_{ik} H_k \]

where \( G_{ij} \) is the cross spectral density between the input \( i \) and the output \( y \), \( G_{ik} \) is the cross spectral density between the inputs \( i \) and \( k \) and \( H_k \) is the transfer function between the input \( j \) and the output.\( G_{ij} \) for the special case of uncorrelated inputs the cross-spectral terms between the individual inputs \( G_{ij} \), \( i \neq j \) are zero and the system may be analysed by measuring the transfer function directly for each input with a dual-channel Fourier analyser. The contribution of the input \( i \) to the total noise measured at the output point \( (G_{ii}) \) will then be given by the following relationship:

\[ (G_{ii}) = |H_i|^2 G_{ii} = Y_i^2 G_{ii} \]

where \( (G_{ij}) \) is the noise at the output measurement point \( y \) due to the input \( i \) and \( G_{ij} \) represents the squared coherence function between input \( i \) and output \( y \).

If the inputs to the system are correlated, then the cross spectral terms between the inputs are no longer zero and the transfer function calculated directly with a dual channel analyser will include part of the effects of the other sources. A multi-channel acquisition system ensures, by using the full expression (1), an accurate estimation of the transfer function. This value may then be used in equation (2) to obtain the contribution of input \( i \) to the noise level at the output measurement point if all the other inputs were suppressed.

Contamination from other sources of the autospectral density \( G_{ii} \) used in this expression will normally lead to an overestimation of the contribution of the source. A better estimate will then be obtained by conditioning out the spectral components that are coherent with the other inputs in order to calculate the part of the contribution of input \( i \) to the output that is completely independent of the other inputs. The independent contribution of the input \( i \) to the output spectrum \( G_{ii} \) is then given by the expression [1]:

\[ G_{ii}(\omega) = G_{ii}(\omega) - \sum_{j \neq i} G_{ij}(\omega) \]

where the suffix \((i-j)\) indicates that the effects of the \((i-j)\) remaining inputs have been conditioned out of the original autospectral terms. The partial coherence function \( \eta_{ij}(\omega) \) will be similar to the ordinary coherence function \( \eta_{ij}(\omega) \) if the input \( i \) is the only significant source. In the case of highly coherent sources the partial coherence can become extremely small and the precision of the estimation of the transfer function is reduced accordingly. Providing the inputs are not too coherent and a sufficient number of averages is obtained the expression (3) fixes a lower limit for the contribution of input \( i \) to the output noise level.

PRACTICAL APPLICATION

These techniques have been applied previously by the authors to the inlet valves of a high pressure steam turbine and gave results which corresponded well with those obtained using sound intensity measurements and calculations of the effects of the valves on the noise level at various points [2]. The technique has since been tested on a high power turbo-feed pump in a 1300 MW nuclear power station and a number of experiments have also been carried out in a reverberant room, using loudspeakers. The most important factors that have been examined in these tests are the effects of reverberation and propagation time, the effects due to inter-source contamination, the effects of coherence between the inputs and the choice of reference signals.

In the laboratory tests two identical loudspeakers were placed 2 metres apart with a reception point situated at 2.5 metres from both sources. The reference signals were obtained using microphones placed close to the loudspeaker membrane. The relatively small distances involved meant that the propagation delay had a negligible effect on the coherence and no correction was necessary. It was found that the changes in coherence were relatively small as the record length of the individual samples was increased above an
eighth of the reverberation time (4s). Use of this relatively short duration enabled the frequency range to be maintained at a reasonable level using a limited number of points in each record. Two identical loudspeakers were used for.

The effects of contamination between sources were examined by varying the level of one source relative to the other in a well-defined frequency range and calculating the contribution of each source using the technique of residual spectra to calculate the transfer functions and the corresponding maximum and minimum estimates for the contribution to the output (expressions (2) and (3)). These tests were carried out with signals that were partially coherent over a given frequency range to see at what level of input the weaker source was contaminated significantly by the stronger source.

For differences between the input signals not exceeding 20 dB the inter-source contamination had little effect, even with partially coherent inputs. The calculated value of the maximum contribution of one of the sources obtained with both sources in action is compared with the output spectrum measured when the second source is switched off in figures 1 and 2. It can be seen that the essential features of the source contribution are reproduced, with a certain number of errors due to the propagation path losses and insufficient averaging. In the last configuration, the reference signal of the second source is due entirely to the contamination. If the multiple input model is applied in this case the energy will be assumed to be partially due to this source and the errors can be greater due to the losses in coherence between the inputs and the output. In this particular configuration, the errors were relatively small (figure 3) but this may not always be the case. A comparison of the ordinary and partial coherence of each input with the output will usually reveal which is the weakest source.

The application to a turbo-feed pump necessitated the use of a five input model for the main noise sources. The maximum limit contributions were calculated for each source and compared with each other. The exhaust seemed to be the main source of noise at the output point, especially at low frequencies and for the peaks in the power spectra (figure 4 and 5). Calculations of sound power using vibration measurements also indicated that the exhaust was the most important source.

CONCLUSIONS

The tests that have been carried out indicate that the coherence techniques may be used to estimate the contribution of sources in relatively reverberant conditions. Losses in coherence limit the accuracy of the final result, but do not seem to lead to errors in assessing the relative importance of the sources. These results can be improved upon with the use of several reference signals for each source and longer record lengths associated with an increased number of averages.

REFERENCES


CARACTERISATION ET LOCALISATION DES SOURCES ACoustiques PAR DES METHODES D'OPTIMISATION.

C. Legros - M. Sidki - J.P. Guilhot


INTRODUCTION

Connaissant la pression et l'accélération acoustiques sur une surface Z, la relation de Kirchhoff nous permet de déterminer la pression en tout point M d'une surface Z', extérieure à Z: c'est un problème direct.

Dans un problème inverse, au contraire, on cherche à remonter à l'état vibratoire sur Z à partir de données acoustiques sur Z'. L'intégrale de Kirchhoff n'étant pas inversible, l'unicité de la solution n'est pas assurée. L'identification des sources pourrait se faire par décomposition du champ acoustique sur une base d'ondes appropriée et rétropropagation des fronts d'onde. Le cas d'une antenne plane a été étudié [1]. La pression acoustique en un point source est obtenue par rétropropagation jusqu'à ce point de toutes les ondes sélectionnées.

Cette méthode a des restrictions:
- dans le calcul de la pression au point M, toutes les mesures doivent être prises simultanément pour conserver les phases. Sinon, il faudrait avoir une référence de phase prise absolument sur la source.
- elle ne permet pas de retrouver les coordonnées des sources.

La résolution analytique du problème inverse n'étant pas possible et l'approche par rétropropagation n'identifiant que totalement les sources, nous présentons ici une résolution numérique.

FORMULATION DU PROBLEME INVERSE COMME UN PROBLEME DE MINIMISATION.

Considérons un champ acoustique engendré par n sources caractérisées par leurs 3 coordonnées spatiales et par les 2 composantes des amplitudes complexes des signaux qu'elles émettent. On cherche à obtenir un système de 3n équations à 5n inconnues.

La résolution numérique de ce système se fera en trois étapes:
- hypothèses sur la nature des sources et sur la propagation.
- déduction des expressions analytiques des 3 composantes du vecteur intensité acoustique.
- formulation de la fonction à minimiser:

\[ f(S_1, S_2, \ldots, S_n) = \sum_{i=1}^{n} \left( \left| \mathbf{r}_{i} \cdot \mathbf{I}_{m} - \mathbf{r}_{i} \cdot \mathbf{I}_{e} \right| \right)^{2} \]

ou \( \mathbf{I}_{m} \) et \( \mathbf{I}_{e} \) sont les composantes mesurées et calculées selon x du vecteur intensité au point i.

Principe de la méthode

La fonction \( f(S_1, S_2, \ldots, S_n) \) sera minimisée par la méthode du gradient. Formons un vecteur:

\[ \mathbf{X} = \left[ x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_n, y_n, z_n \right] \]

où \( x_i, y_i, z_i, A_i \) sont les caractéristiques de la source \( S_i \).

A partir d'un vecteur initial \( \mathbf{X}_0 \), on ainsi se donnera, on forme une suite de vecteurs \( \mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_t \) telle que la fonction converge le plus rapidement possible, ce qui est réalisé si deux vecteurs successifs sont tels que:

\[ \mathbf{X}_{t+1} = \mathbf{X}_t - C \frac{\partial f}{\partial \mathbf{X}_t} \]

avec \( C = \frac{2}{\lambda_{\text{min}} - \lambda_{\text{max}}} \)

où \( \lambda_{\text{min}} \) et \( \lambda_{\text{max}} \) sont respectivement la plus petite et la plus grande valeur propre de la matrice formée par les dérivées secondes de \( f(x) \).

Cas particulier: sources ponctuelles en champ libre.

Pour ce cas simple, nous envisagerons trois types de sources, et nous donnerons dans chaque cas les expressions analytiques des trois composantes du vecteur intensité acoustique.

Deux sources monochromatiques ponctuelles \( S_1 \) et \( S_2 \) émettant à la même fréquence \( f \) et de caractéristiques respectives \((x_1, y_1, z_1, A_1(f))\) et \((x_2, y_2, z_2, A_2(f))\) créent en un point \( M \) de coordonnées \((x, y, z)\) une intensité acoustique dont la partie active de la composante selon l'axe x s'écrit (voir figure i):

\[ R_{\text{I}}(x, y, z) = a_1(x, y, z) \cdot \frac{A_1}{pcr_1^2} \cos \frac{\pi x}{\lambda} \left( \frac{\lambda}{pcr_1} \right) \]

\[ + \frac{A_2}{pcr_2^2} \left( \frac{\lambda}{pcr_2} \right) \cos \frac{\pi x}{\lambda} \left( \frac{\lambda}{pcr_2} \right) \]

\[ + \frac{\sin \frac{\pi x}{\lambda} \left( \frac{\lambda}{pcr_1} \right)}{kr_1 \left( \frac{\lambda}{pcr_1} \right)} \left( \frac{\lambda}{kr_1} \right) \]

avec \( kr_1 = k(r_2 - r_1) \).

Elle comporte, en plus des intensités dues séparément à chacune des sources, deux termes exprimant l'interaction entre les deux sources.

Pour deux sources de même bande, l'expression de l'intensité active selon l'axe \( x \) découle de l'expression (2) en intégrant chacun des termes sur toute la bande fréquentielle d'analyse. L'expression de l'intensité active en un point \( M \) d'un champ sonore créé par \( n \) sources de bande décortillées s'obtient en se gardant, dans l'expression précédente, que les deux premiers termes.

DIMENSION DE LA SURFACE DE MESURE ET PAS D'ECHANTILLONNAGE SPATIAL DE L'INTENSITE ACoustique.

Détériorons les conditions de mesure qui permettent la séparation des sources et donc leur identification par retour inverse. Nous distinguons les 3 types de sources décrites ci-dessus et nous choisissons, comme surface de mesure, un plan horizontal.

Dans le cas général, on montre (2) que ce source monochromatiques seront séparables si l'on choisit une ouverture angulaire 2 \( \theta_{\text{max}} \) telle que:

\[ \sin \theta_{\text{max}} \leq \frac{\lambda}{2 \Delta \text{min}} \]

où \( \Delta \) est la plus petite distance séparant deux sources.

En champ lointain, le pas d'échantillonnage spatial de l'intensité doit être tel que:

\[ \sin \theta_{\text{max}} \leq \frac{\lambda}{2 \Delta \text{max}} \]

où \( \Delta \) est le nombre de sources.

En champ très proche des sources, toute l'information sur l'intensité acoustique se trouve dans une fenêtre de largeur environ égale à la distance entre les deux sources extrêmes. Nous adoptons une couverture de mesure de largeur comprise entre celle définie pour le champ lointain et une fois la distance entre les deux sources extrêmes.
Quant au pas d'échantillonnage de l'intensité acoustique, on gardera les mêmes règles que pour le champ lointain. En effet, dans le champ proche l'hypothèse $k r < 1$ est encore vérifiée.

Pour des sources larges/bandes corrélées, l'ouverture angulaire nécessaire à la séparation en champ lointain est donnée par (3), en prenant $A = A_{n0}$. Dans le champ proche, l'échantillonnage sera correct si l'on choisit les points de mesure tels que: $|r'_i - r_0|, |r'_i - r_0| < \frac{\lambda}{n \Delta \lambda_{n0}}$

**SIMULATION ET PRESENTATION DES RESULTATS.**

Nous avons développé, sur un calculateur HP 9000, un programme simulant la résolution du problème inverse par la méthode d'optimisation décrite. Nous noterons $\sigma_{\text{opt}}$ l'écart-type réduit des caractéristiques des sources:

$$\sigma_{\text{opt}} = \frac{1}{2m} \left\{ \frac{\sum (x_{i_{1k}} - x_{i_{2k}})^2}{2x_{i_{1k}}^2 + x_{i_{2k}}^2} + \frac{\sum (x_{i_{3k}} - x_{i_{4k}})^2}{2x_{i_{3k}}^2 + x_{i_{4k}}^2} + \frac{\sum (x_{i_{5k}} - x_{i_{6k}})^2}{2x_{i_{5k}}^2 + x_{i_{6k}}^2} \right\}^{\frac{1}{2}}$$

où $x_{i_{1k}}, x_{i_{2k}}, x_{i_{3k}}, x_{i_{4k}}, x_{i_{5k}}, x_{i_{6k}}$ représentent les valeurs exactes et $x_{\text{opt}}$ l'écart-type réduit de l'intensité.

Les résultats de la simulation sont donnés dans le tableau 1, dans le cas du champ proche, et nous noterons:

- **Couv**: la couverture spatiale des sources.
- **Pech**: le pas d'échantillonnage considéré.
- **f**: l'unité des amplitudes des signaux émis par les sources.

**Cas de 2 sources monochromatiques de fréquence f.**

On constate que pour un nombre de points de mesure relativement faible (inférieur à 3), on a pu retrouver, avec une bonne précision ($\sigma < 4.10^{-3}$), les positions des sources, ainsi que les amplitudes des signaux émis.

**Cas de 2 sources large bande corrélées (bruit blanc).**

Trois points de mesure suffisent pour retrouver les coordonnées des sources ainsi que leur énergie dans l'octave 250Hz.

**Cas de 2 sources large bande décroîtées.**

Les résultats montrent qu'il faudra se placer à une hauteur inférieure à deux fois l'écartement entre les sources. Au-delà de cette hauteur, la séparation devient très difficile et il faudra prendre une couverture spatiale importante pour avoir suffisamment de variations d'intensité. D'autre part, la convergence devient très lente et par conséquent le nombre d'itérations est important.

**CONCLUSION.**

Les résultats obtenus pour les 3 types de sources considérées montrent qu'en respectant les conditions relatives à la dimension de la surface de mesure et au pas d'échantillonnage spatial de l'intensité acoustique, on a pu localiser et caractériser des sources ponctuelles.

Pour trouver le modèle représentant le mieux la source de bruit étudiée, on pourra procéder de la manière suivante:

1. Partir d'un modèle constitué de deux sources ponctuelles; si, au bout d'un nombre d'itérations donné, la fonction $f(x)$ n'est pas minimisée, le modèle à deux sources ne sera pas convenable.
2. Augmenter progressivement le nombre de sources ponctuelles du modèle, jusqu'à ce que la fonction $f(x)$ soit minimisée.

4. Partir du modèle final trouvé, prévoir toute grandeur acoustique relative au champ sonore créé par la source de bruit considérée.

**BIBLIOGRAPHIE.**


**Figure 1 - Cas de 2 sources ponctuelles.**

**Tableau 1 - Résultats**

<table>
<thead>
<tr>
<th>$X_{ex} = (2m, 2u, 4m, 2u)$</th>
<th>$X_{ex} = (2.5m, 3u, 3.5m, 3u)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Z = 2m$</td>
<td>$\sigma_{\text{opt}} = 0.05$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Couv = 1m</th>
<th>$P_{\text{ech}} = 0.5m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sources</td>
<td>$X_{40} = (1.98m, 1.99u, 4.01m, 1.99u)$</td>
</tr>
<tr>
<td>monochromatiques $\sigma = 4.10^{-3}$</td>
<td></td>
</tr>
<tr>
<td>sources large</td>
<td>$Y_{13} = (1.93m, 2.11u, 4.06m, 2.11u)$</td>
</tr>
<tr>
<td>bande corrélées $\sigma = 0.04$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Couv = 6m</th>
<th>$P_{\text{ech}} = 2m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sources large</td>
<td>$X_{20} = (2.19m, 1.96u, 3.80m, 1.96u)$</td>
</tr>
<tr>
<td>bande décorrélées $\sigma = 0.06$</td>
<td></td>
</tr>
</tbody>
</table>
SOUND POWER LEVELS OF NOISE SOURCES DETERMINED BY VARIOUS KINDS OF MEASURING METHODS

Hiroo Yano, Hideki Tachibana and Masaru Koyasu

Institute of Industrial Science, University of Tokyo, 7-22-1 Hoppangi, Minato-ku, Tokyo 106, Japan

INTRODUCTION

Concerning the sound power level measurement of various noise sources, a set of Japanese Industrial Standards (JIS) are now being prepared, referring to the ISO standards (ISO 3740 series and the Drafting work regarding the sound intensity method).

As a basic study for this work, the sound power levels of several sound sources were measured by various methods based on the ordinary p-square method and the sound intensity method, and these measuring methods were compared. (Based on this experimental study, inter-laboratory tests regarding the sound power level measurements are being performed in Japan.)

METHOD FOR SOUND POWER LEVEL MEASUREMENTS

The various methods for measuring the sound power level are classified into three kinds according to the measuring principles, as follows:

1) p-method(1) (Free or semi-free field method)

By measuring the sound pressure (p) on a measurement surface surrounding the sound source, the sound intensity (I) is approximately from it according to the relation of \( I = p^2 / \rho c \). Then, by integrating the sound intensity (or area weighted summation) over the surface, the sound power is obtained. (ISO 3744, 3745 and 3746 are based on this method.)

2) p-method(2) (Diffuse sound field method)

By measuring the sound pressure in a diffuse sound field (reverberant room) in which the sound source is located, the sound energy density (E) of the field is approximated from it according to the relation of \( E = \rho c / p^2 \). Then, from the mean value of the sound energy density and the sound absorption area of the field, the sound power of the source is obtained. (ISO 3741, 3742 and 3743 are based on this method.)

3) I-method (Sound intensity method)

By measuring the sound intensity on an arbitrary measurement surface surrounding the sound source directly by the sound intensity technique, and by integrating it over the surface, the sound power of the source is obtained. (ISO standard based on this method is now being prepared.)

Regarding the Japanese Industrial Standards, the precision method based on the method 1) has been instituted (JIS Z 8752), and the Engineering and Survey methods based on the method 2) have been drafted (fiscal 1986). In fiscal 1986, the precision method based on the method 2) is now being prepared.

In this study, these measuring methods were compared by measuring the sound power levels of four kinds of sound sources, as follows.

MEASURING FACILITIES AND INSTRUMENTATION

For the measurements under free field condition, an anechoic room of 70m³ air volume which complies with the requirements stated in ISO 3745 was used. When making it semi-free field condition, plywood boards of 50mm thickness and vinyl chloride boards of 10mm thickness were laid over the floor.

For the measurements under diffuse sound field condition, a reverberation room of 200m³ air volume which complies with the requirements stated in ISO 3741 was used. Its reverberation time is about 7 seconds in 500Hz.

For the sound pressure level measurements, a 1/2 in. condenser microphone (B&K 4133, free field type) and a measuring amplifier (B&K 2134) were used.

In the case of diffuse field measurements, the random incident correction was made for the sensitivity of the microphone.

For the sound intensity measurements, B&K 3360 sound intensity measuring system including two 1/2 in. condenser microphones (d=12mm) was used.

MEASURED RESULTS

(A) A Loudspeaker Source

First, a flat type loudspeaker (Technics SR-1000) was used and an artificial broad band noise (M-sequence signal) was radiated. The measurements were made by (a) the p-method (1) in the reverberation room (after ISO 3741), (b) the p-method (1) in the semi-anechoic room (after ISO 3745) in which 10 measuring points were chosen on a hemispherical surface of 1m radius (see Fig.-1), (c) the I-method in the semi-anechoic room by using the same measurement surface and points, and (d) the I-method in the semi-anechoic room by using a rectangular parallelepiped surface of 35cm×35cm×10cm which were divided into 77 segments.

The comparisons are shown in Fig.-2, and it can be said that the measured results are markedly in good agreement except for the high frequencies where the differences of about 2dB at most are seen.
(D) A Mini-Compressor

As an example of actual noise source, a mini-compressor (IWATA OLP-02P 0.2kW) was chosen and its sound power level was measured. Fig. 6 shows the comparisons of results measured by (a) the p-method(1) in the semi-anechoic condition, (b) the p-method(2) in the reverberant condition and (c) the I-method in the semi-anechoic condition using a rectangular parallelepiped measurement surface of 50cm x 40cm x 60cm which was divided into 384 segments. In these results, the A-weighted sound power levels are in good agreement, but relatively large differences are seen in each 1/3 octave bands, because this source has discrete spectral components.

In order to examine the effects of sound field condition in the case when the I-method is applied, three kinds of measurements were performed for this source in the semi-anechoic room, an ordinary office room by measuring 6.3m by 6.6m by 3m (reverberation time is 0.7 seconds in 500Hz) and the reverberation room. As shown in Fig. 7, fairly good agreements can be seen in spite of extreme differences of the sound field. This might indicate the efficiency of the sound intensity method in the sound power level measurements.

(B) A Loudspeaker Type Reference Sound Source

As the second sound source, B&K 4205 reference sound source (loudspeaker type) was used. First, the measurements were made by (a) the p-method(1) in the anechoic condition (after ISO 3745), (b) the p-method(1) in the semi-anechoic condition (after ISO 3745) (c) the p-method(2) in the reverberant condition (after ISO 3741). As shown in Fig. 3, it can be seen that the values measured in the anechoic and semi-anechoic conditions are markedly in good agreement in middle and high frequencies, but systematic differences are seen in low frequencies. This might be caused by the difference of with and without reflective plane. On the other hand, the results measured in the semi-anechoic condition and reverberant condition are slightly different in high frequencies.

Next, the measurements by the I-method were performed in the semi-anechoic condition by using three kinds of measuring surfaces: a rectangular parallelepiped surface of 50cm x 50cm x 50cm which was divided into 500 segments, another rectangular one of 50cm x 30cm x 35cm which was divided into 204 segments, and a cylindrical surface of 20cm radius and 40cm height. In the last case, the automatic scanning technique was applied. As shown in Fig. 4, those three kinds of results are in good agreement, but they are slightly lower than the values measured by the p-method(1).

(C) A Fan Type Reference Sound Source

As an aerodynamic sound source, ILC reference sound source was used. The measurements were made by the p-method(1) in the semi-anechoic condition (after ISO 3745) and the p-method(2) in the reverberant condition (after ISO 3741), and a sufficiently good agreement was obtained as shown in Fig. 5. Concerning this source, the I-method was not applied because of the presence of wind.
MEASUREMENT OF ACOUSTIC ENERGY EMITTED BY IMPULSIVE SOUND SOURCES

Hideki Tachibana and Hiroyo Iono

Institute of Industrial Science, University of Tokyo, 7-22-1 Roppongi, Minato-ku, Tokyo 106, Japan

INTRODUCTION

Concerning stationary sound sources, the sound power level is clearly defined, and its measuring methods have been established. Concerning transient sound sources such as impulsive sources, however, the way of defining and measuring their acoustic outputs have not been unified.

In this paper, the sound energy flux level which represents the total sound energy emitted by an impulsive sound source is defined, and its measuring methods using both of the p-square method and sound intensity method are studied.

DEFINITION OF SOUND ENERGY FLUX LEVEL

The sound power level of a stationary source is defined as:

$$L_W = 10 \log \frac{W}{W_0}$$  \hspace{1cm} (1)

where, $W$ is the sound power of the source (W), $W_0$ is the reference sound power ($10^{-12}$ W).

Concerning an impulsive sound source, it seems reasonable to define the following quantity; the sound energy flux level.

$$L_E = 10 \log \frac{E}{E_0}$$  \hspace{1cm} (2)

where, $E$ is the total sound energy (sound energy flux) emitted by the source, and $E_0$ is the reference acoustic energy ($10^{-12}$ J).

MEASURING METHODS

There are three different methods for the measurement of sound power level of stationary sound sources, that is the p-square method in free sound field and that in diffuse sound field and the sound intensity method. Concerning transient sound sources, there can be similar three ways for the measurement of sound energy, as follows.

(1) Free Sound Field Method

The instantaneous sound intensity $I(t)$ at a point on a spherical surface surrounding a sound source is:

$$I(t) = p(t) \cdot \vec{u}(t)$$  \hspace{1cm} (3)

where, $p(t)$ is the sound pressure and $\vec{u}(t)$ is the particle velocity. In the far field from the source:

$$u(t) = \rho c p(t)$$  \hspace{1cm} (4)

where, $\rho c$ is the characteristic impedance of air.

Accordingly, the total energy flux through unit area (the sound energy flux density) $\sigma$ is expressed as:

$$\sigma = \int_0^\infty I(t) \, dt = \frac{1}{\rho c} \int_0^\infty p^2(t) \, dt$$  \hspace{1cm} (5)

In order to describe this quantity in decibel, the following level (the sound energy flux density level) is defined.

$$L_\sigma = 10 \log \frac{\sigma}{\sigma_0}$$  \hspace{1cm} (6)

Here, the reference value $\sigma_0$ is:

$$\sigma_0 = \frac{20}{S_0} = 10^{-12} \frac{1}{\rho c} \frac{S_0}{T_0} = 10^{-12} (J/m^2)$$

where, $S_0 = 10^{-12}$ W, $T_0$ is 1 s and $\sigma_0$ is 1 m^2.

Accordingly, Eq. (6) can be written as:

$$L_\sigma = 10 \log \left[ \frac{1}{T_0} \int_0^\infty p^2(t) \, dt \right] = L_{p\sigma}$$  \hspace{1cm} (7)

where, $L_{p\sigma}$ is so called "single event sound exposure level".

The total sound energy flux emitted by the sound source is:

$$E = \int_S \sigma \, dS = \sigma_0 \cdot 4\pi r^2$$  \hspace{1cm} (8)

where, $S$ means the spherical surface, $\sigma$ is the mean value of $\sigma$ over the surface and $r$ is the radius of the sphere. Accordingly, the sound energy flux level can be written as:

$$L_E = 10 \log \frac{\sigma_0 \cdot 4\pi r^2}{E_0}$$

$$= L_\sigma + 10 \log \frac{n^2}{\rho c} + 11 = L_{pE} + 10 \log \frac{n^2}{\rho c} + 11$$  \hspace{1cm} (9)

where, $L_{pE}$ is the energy mean value of $L_{p\sigma}$ over the measuring surface and $S$ is im. (In the case when the sound source is located on a reflective plane, 4m^2 and 11 in Eq. (9) must be substituted by 2m^2 and 8, respectively.)

(2) Diffuse Sound Field Method

When a transient sound source is located in a diffuse sound field (reverberation room), the following equation should hold:

$$E(t) = V \frac{dE_g(t)}{dt} + \frac{c}{4} \frac{E_g(t)}{A}$$  \hspace{1cm} (10)

where, $R(t)$ is the instantaneous sound energy flux emitted by the source, $E_g(t)$ is the sound energy density in the field ($W/m^3$), $V$ is the air volume of the reverberation room ($m^3$) and $A$ is the equivalent sound absorption area ($m^2$).

By making $t \to \infty$, $E_g(t)$ becomes zero, and so, by integrating Eq. (10), the sound energy flux can be expressed as:

$$E = \int_0^\infty E(t) \, dt = \frac{c}{4} \int_0^\infty \frac{E_g(t)}{A} \, dt$$  \hspace{1cm} (11)

Here, in the diffuse sound field, the following relation should hold:

$$E_g(t) = \frac{p^2(t)}{\rho c T}$$  \hspace{1cm} (12)

Hence:

$$E = \frac{A}{4 \rho c} \int_0^\infty p^2(t) \, dt$$  \hspace{1cm} (13)

and

$$L_E = L_{pE} + 10 \log \frac{A}{S_0} - 6$$  \hspace{1cm} (14)

That is, $L_E$ can be obtained by measuring $L_{pE}$ in the reverberation room (in actual measurements, spatial mean value should be obtained) and the reverberation time.

(3) Sound Intensity Method

The third method is to measure the sound intensity directly by using the sound intensity measuring technique. That is, the sound energy flux density $\sigma$ is obtained by integrating the instantaneous sound intensity on the measuring surface surrounding the sound source, and by integrating its value over the surface (or area-weighted summation), the sound energy flux $E$ can be obtained.


EXPERIMENTAL STUDIES

(1) Study on Measurement of Sound Energy Flux Density by the Sound Intensity Method

First, a basic study was made in order to examine the measurement of the sound energy flux density level \( L_E \) by the 2-microphone sound intensity technique. For the sound source, a condenser type loudspeaker was located in an anechoic room, and 1kHz tone bursts of a constant amplitude and various steps of duration times were radiated from it. At a point of 2.5m from the source, \( L_{pE} \) and \( L_E \) were measured according to the method (1) and (3), respectively. In this measurement, B&K 3360 sound intensity measuring system using two 1/2 in. condenser microphones (\( d=12 \text{mm} \)) was used.

Fig.-1 shows the measured results. In this figure, 0dB is the sound pressure level or the sound intensity level when the source signal was stationary. In the result, it can be seen that \( L_{pE} \) and \( L_E \) are markedly in good agreement as expressed by Eq.(8), and their antilog values are accurately proportional to the duration time of the source signal.

(2) An Example of Sound Energy Flux Level Measurement by the Three Methods

According to the three methods, \( L_E \) of an artificial impulsive sound source was measured as follows.

For the sound source, a flat type loudspeaker was used, and two kinds of noise burst signals were radiated from it: one was a sequence of \( M \)-sequence signal of 100ms duration time and another was that of 10ms duration time. For the measuring instruments, the system mentioned above was used.

First, the measurement was performed in an anechoic room by locating the source on a reflective panel as shown in Fig.-2. Ten measuring points were chosen on a hemispherical surface of 1m radius according to ISO 3745, and \( L_E \) in 1/3 octave bands at these points were measured by both of the \( p \)-square method and the sound intensity method. From these results, the sound energy flux levels \( L_E \) were calculated.

Besides these measurements, a rectangular surface (35cm x 35cm x 10cm) was set in the near field of the sound source, and the measurement by the sound intensity method was made. In this case, the measuring surface was divided into 77 areas.

Next, the sound source was located in a reverberation room of 200m\(^3\) air volume, and \( L_{pE} \) values were measured at five points by the \( p \)-square method. From the mean value of the measured \( L_{pE} \) and the reverberation time of the room, \( L_E \) values were calculated by Eq.(14).

The experimental results for the two kinds of noise burst signals are plotted in Fig.-3 and 4. If examined these results in detail, it may be seen that the values measured in the reverberation room are slightly larger than others and those measured by the sound intensity method in the near field are relatively low, but, roughly speaking, \( L_E \) values measured in the four different ways are fairly in good agreement with 2 or 3dB differences.

REFERENCES

TRAITEMENT SPATIAL DES DONNES OU DES IMAGES EN IMAGE ACOUTIQUE AU CHAMP PROJET.

B. Seguet. P. Wetta.
Metraffr B.R.S. 64, Chemin des Mouilles B.P. 182 69132 Écully Cedex - France

INTRODUCTION

Cette contribution s'inscrit dans le problème général d'analyse de bruits industriels - constitués plus souvent d'une combinaison acoustique des champs acoustiques, au moyen des techniques de traitement des données issues de l'acoustique. Dans ce contexte, il est nécessaire de considérer la diversité des sources de l'émission sonore. Différentes approches du problème existent et se répartissent en diverses techniques de traitement pour l'identification des sources de l'émission sonore. Les résultats obtenus concernent des diverses techniques que l'on peut rencontrer, aussi bien dans les structures de nombreux auteurs, dans les circuits techniques de champ et de son [3].

1. Le traitement intensimétrique conduit principalement à la quantification du bruit total émis.

La protection du champ acoustique sur une base sonore [2] permet de rétropropager les composantes obtenues vers le bruititre et de reconstituer ainsi une image du bruit au plus près des sources. L'examen des images conduit à mettre en évidence les différences de répartition de l'émission et à définir le moyen de détecter les sources de bruit.

2. L'objectif de cette contribution est d'analyser, de mettre en évidence des résultats expérimentaux et de comparer l'aptitude de ces deux méthodes à identifier et localiser des sources de bruit plus ou moins complexes.

I - TRAITEMENT INTENSIMÉTRIQUE

Nous considérons un bruitqueur acoustique et une surface fermée entourant ce bruitqueur (figure 1). L'analyse des champs acoustiques est effectuée par un réseau de capteurs répartis sur cette surface. En chaque point du magage, nous mesurons l'intensité acoustique sous la base de la source de bruit.

\[ T_n = \frac{1}{T} \int_0^T \nu(t) \, dt \]

ou \( T_n \) et \( \nu(n) \) représentent respectivement la pression acoustique et la vitesse partielle normale en chaque point du magage. L'intensité acoustique représente le flux d'énergie à un point de mesure considéré.

Notons que cette propriété reste valable même en présence de bruitleurs extérieurs à la surface.

L'analyse des cartographies d'intensité nous permet également de mettre en évidence certaines caractéristiques du bruit émis. On peut en particulier obtenir la puissance rayonnée par une source sous l'effet d'un bruitleur complet, à condition que ceux-ci soient spatialement suffisamment dissociés.

Toutefois, ne donnant un énergie point à point, ne permet pas de déterminer les surfaces d'ondes de la source et de localiser les sources de bruit au sens strict du terme.

Pour illustrer, on montre ainsi des résultats de simulations du calcul intensimétrique effectuées dans le cas d'une plaque carrée (Côle = 50 cm) vibrante sur un moteur. Le seuil d'énergie est fixé à 234.24 Hz. (Fréquence critique de la plaque = 2490 Hz ; on choisit k = 7.7 ; L = demi-environ du magage)

On peut montrer théoriquement qu'un tel champ de vibration (figures 1 et 2) conduit par effet de compensation des sources d'accélération, à un rayonnement en champ lointain. Dans ce cas, le calcul de l'intensité est conduit sur une surface plane carrée de 2 cm sur 2 cm de la plaque.

Les résultats montrent (figures 3) :

- des zones d'intensité très forte en regard des quatre coins de la plaque.
- des zones de forte intensité sont également présentes sur tout le pourtour de la plaque.
- les résultats obtenus en cette technique pour identifier des sources de l'émission sonore. Ce phénomène de distribution de sources est en parfait accord avec les résultats obtenus dans le cadre de la répartition spatiale du champ. La cohérence spatiale et temporelle du champ acoustique est également présente sur toute la surface de la plaque.

II - DECOMPOSITION DES CHAMPS ACOUSTIQUES SUR UNE BASE D'ONDES PLAINES

Cette méthode est largement illustrée dans la littérature [3] ou [4].


Dans le cas de la radiation acoustique du plan P1 vers le plan P2, les résultats obtenus en cette technique pour identifier des sources de l'émission sonore. Ce phénomène de distribution de sources est en parfait accord avec les résultats obtenus dans le cadre de la répartition spatiale du champ. La cohérence spatiale et temporelle du champ acoustique est également présente sur toute la surface de la plaque.

- des zones d'intensité très forte en regard des quatre coins de la plaque.
- des zones de forte intensité sont également présentes sur tout le pourtour de la plaque.
- les résultats obtenus en cette technique pour identifier des sources de l'émission sonore. Ce phénomène de distribution de sources est en parfait accord avec les résultats obtenus dans le cadre de la répartition spatiale du champ. La cohérence spatiale et temporelle du champ acoustique est également présente sur toute la surface de la plaque.
II.2. - Application de la méthode à un cas expérimental

Le principe développé dans le cas d'une géométrie plane, a été transposé à d'autres types de géométrie. Le cas traité expérimentalement est celui du bruit émis par un cylindre soumis à une excitation ponctuelle (figure h). L'expérimentation a été conduite avec une antenne acoustique plane. Par des rotations de la lampe, on a obtenu des paquets d'ondes rayonnées dans des directions autour du cylindre ont été mesurés avec un pas \( \Delta x, \Delta y \) (quelques centaines de points de mesure).

Sur les figures 7a et 7b, nous donnons les distributions de sources obtenues en module, dans le cas d'une source acoustique placée devant le cylindre puis d'une excitation vibratoire ponctuelle. L'examen de ces images appelle les remarques suivantes :

- l'exemple de la source acoustique met en évidence la fonction d'appareil de l'imageur dans le cas de la géométrie étudiée (figure 7b). Il s'agit d'une fonction oscillante rapidement amortie.

- la distribution obtenue pour l'excitation vibratoire (figure 7a) montre une source concentrée au point d'excitation (pointonnage local) et des ondes se développant sur tout le cylindre, principalement vers les bords.

II.3. - Séparation des composantes localisées et étendue du champ rétropropagé

L'analyse de l'image précédente conduit à vouloir séparer en deux composantes : une réponse locale \( s_{loc}(x) \) et une réponse étendue \( s_{etz}(x) \).

L'objet de ce dernier paragraphe est de quantifier la part énergétique rayonnée par chaque composante.

On sait que la fonction \( h(X, Y) \) pour une source ponctuelle située en \( X_0, Y_0 \) peut s'écrire \( s_{loc}(x) = c_0 h(X_0, Y_0) \), \( c_0 \) nombre complexe. Séparer les réponses locales et étendues revient dans ce cas à estimer \( c_0 \). Pour cela, on peut se donner les contraintes suivantes :

* au point d'excitation, on peut écrire :
  \[ s_{etz}(X_0) = c_0 \ h(X_0, Y_0) + s_{etz}(X_0) \]

d'où \( c_0 = s_{etz}(X_0)/h(X_0, Y_0) - s_{etz}(X_0) \).

Les équations précédentes nous montrent que le module du champ étendu est faible vis-à-vis de celui du champ local. On ne réalisera donc \( c_0 \) que dans un domaine limité autour de \( X_0, Y_0 \).

On définit la fonction \( f(x) = s_{etz}(x) = c \ h(x, y) \).

L'intégrale de la fonction sur toute l'aire de l'imageur est donc égale à 1. On utilise alors la méthode de calcul de la réponse locale du type \( h(x, y) \) autour d'un point autre que \( X_0, Y_0 \). On doit donc minimiser le produit d'intercorrélation entre \( f, h \).

On a ainsi séparé l'image rétropropagée (figure 7b) en une composante locale (figure 7c) et une composante étendue (figure 7d). Cette séparation nous permet alors de quantifier indépendamment chaque composante. Les calculs attribuent environ 50 % de l'énergie à chaque composante.

BIBLIOGRAPHIE :


EXPERIMENTAL ANALYSIS OF WHEEL/RAIL NOISE BY NEARFIELD ACOUSTICAL IMAGING

P. Wetta, B. Bégouet

Métravib R.D.S., P.B. 182, 69132 Ecully Cedex, France

INTRODUCTION

The aim of this paper is to present the test results of a study concerning the noise of railway wagons, conducted in collaboration with the S.N.C.F. (Société Nationale des Chemins de Fer Français) in the framework of a research program of the O.R.E. (Office for Research & Experiments of the International Union of Railways) [1].

The topic of this paper is to determine the noise sources present on the rolling mechanism of a wagon running on a network of a straight track. For this purpose it is necessary:

- To use a technique of acoustical measurements, suitable to localize the noise sources with the best accuracy. According to the required space resolution, a nearfield acoustical technique (additive nearfield microphone arrays mounted on the vehicle) was then chosen and a detailed description is given in the next section of this paper.

- To carry out the tests under the most reproducible measuring conditions, since the acquisition of the numerous signals needed for the localization is not simultaneously obtainable during a single train pass.

I. ACOUSTICAL METHOD TO LOCALIZE SOURCES

The space analysis of the nearfield acoustical field is usually carried out by mapping a measurement surface surrounding the noise source, by means of a microphone array, located in the nearfield of the sources.

Wheel/rail noise is not stationary and leads to the use of a technique called "Local Analysis" [2]. It consists in using arrays of smaller size than the source. Practically, in order to map the whole wheel, 12 adjacent positions of a 0.28 m x 0.28 m array were defined (figure 1). This array is made up of 4 x 4 microphones. The data acquisition for the various array positions was performed on the same track section to ensure good reproducibility (figure 1).

The recommended processing is then an energy processing for each small array:

- decomposition of the measured field into plane waves;
- calculation of the measured acoustical energy by means of the small array; this energy can be defined as being the sum of energies obtained for each plane component detected by the array.

It is shown that, when the array is located at a distance from the sources less than D^2/λ (D = array size; λ = wavelength), the energy measured in this manner represents the acoustical energy radiated by the sources zone in front of the array, inasmuch as the latter detects the sources included in a volume element subtended by the array outline.

The process repeated for each small array, allows us to obtain an energy mapping of sources (by octave band or other, if required) in the case of a mapping with a large number of points and for non-stationary signals.

II. ANALYSIS OF RESULTS

The analysis of results can be divided into several stages:

II.1. Frequency analysis of signals

The aim of this analysis is to study the spectral distribution of the emitted noise and to determine if it is of the broadband type or is constituted of peaks. Figure 2 illustrates spectrum of the signal given by the array located at position 10 in logarithmic frequency scale.

The analysis of this representation leads to the following comments, namely the spectrum of the emitted noise consists of:

- an important continuous background, which is slightly decreasing in the low frequency range (0 - 400 Hz), then highly decreasing from the 400 Hz cut-off frequency;
- a peaks spectrum emerging from the continuous background mainly in the high frequency range (1 000 Hz) with a peak in the low frequency range (280 Hz). It is shown that the location of these peaks is constant, whatever the running speed, which leads to consider a modal radiation of some elements.

II.2. Time analysis of signals

The goal of the analysis of the signal evolution while rolling over a marked track section is to detect the obvious periodicities (related to a wheel rotation or to a passage on the ties), and to study if some specific track sections are associated with acoustical emissions more significant than others.

This analysis is carried out either according to the time signal of acoustical pressure itself, or according to the time history of the emitted energy by octave bands.

The study of the energetic time histories (figure 3) makes it possible to illustrate the significant level increase, particularly in the 250 Hz octave band, for higher octave bands the emitted energy shows a moderately fluctuating evolution.

II.3. Space analysis

This analysis is carried out according to the local analysis procedure mentioned in section 1 and its aim is to determine if the sound emission, for a given octave band, is associated with the wheel, the rail or a zone near the wheel/rail contact.

The energy maps per octave band (octave 1 kHz, 2 kHz) are given in figure 4 (scaled in maximum percentage). The following conclusions can be drawn:

- For low frequencies (octave 250 Hz) the levels are highest at the center of the wheel.
- For midfrequencies (octaves 500 Hz and 1 kHz) the levels are highest near the rail (and even near the wheel/rail contact for the octave 1 kHz), then they decrease as the position of the array gets higher with respect to the rail.
- For high frequencies (octave 2 kHz) all positions of the array located in front of the wheel exhibit nearly similar high levels.

The arrays mapping areas around the wheel present lower levels.

III. CONCLUSIONS

The set of previous results allows 3 types of radiation mechanisms to be identified, whereas all can be possibly linked to the same noise generating physical mechanism.

At low frequencies (octave 250 Hz)

The sound emission mainly occurs when rolling over weldings between rails. The emission occurs only on a peak (500 Hz), which is likely a wheel mode (high level in the center of the wheel).

It may be noted that an artificial excitation, described in a previous study, allows a peak at 370 Hz to be shown during an axial shock produced on the wheel.

At midfrequencies (500 Hz, 1 kHz)

The noise emission is prominent in a zone in the neighborhood of the contact area. Moreover the spectrum
general shape is that of a low-pass filter (cut-off frequency about 400 Hz).

All these remarks lead to consider that these features agree with the Remington model. From this formulation (\[3\]), the main excitation is assumed to be caused by the combined spectra of the roughnesses of the wheel and rail, which are then filtered by various operators characterizing the various mechanical and acoustical transfers:

- Low-pass filtering due to the extent of the contact area (Hertz ellipse).
- Mechanical transfer function (impedance, local stiffnesses as seen from the contact area, modal response).
- Acoustical mechanical transfer function (efficiency factor of the radiation).

Therefore the general shape of the spectrum (low-pass filtering) seems to be obtained through the filtering by the Hertz ellipse, which is due to the fact that the wheel/rail contact is not punctual but is distributed on a certain ellipse-shaped surface.

For the case considered, as the cut-off frequency is about 400 Hz at a speed of 60 km/h, the relation, given by Remington, provides a value of 7 mm for the characteristic size of the ellipse \(\sqrt{pq}\), which is of the order of magnitude of the contact size.

At high frequencies (octave: 2 kHz)

The noise emission is determined by the acoustical radiation of the wheel modes. The results show that the modes of the whole wheel are concerned, since the level is nearly constant on its whole surface.

With regard to the generating mechanisms it is however not possible to determine if the excitation is due to the roughness of both wheel and rail or other phenomena, like slidings.

BIBLIOGRAPHY


COMMUNITY NOISE SURVEYS IN AUSTRALIA

Warren D. Renew
Division of Noise Abatement, 64-70 Mary Street,
Brisbane, Queensland, 4000, Australia

INTRODUCTION

Australian noise surveys have generally followed the model provided in the USA [1]. Surveys have been carried out in the State capital cities of Brisbane [2,3] Melbourne [4], Sydney [5] and Adelaide [6]. They were designed basically to determine the noise climates in several land-use areas as an aid to establishing acceptable noise levels for planning purposes.

SURVEY METHOD AND RESULTS

Site Rating

Australian Standard AS1055 [7], which has been the basis for all the surveys, specifies the following standard time periods in a table of estimated average background noise levels:

Morning (6.00 a.m. to 1.00 a.m.),
Day (7.00 a.m. to 6.00 p.m.),
Evening (6.00 p.m. to 10.00 p.m.), and
Night (10.00 p.m. to 6.00 a.m.).

There are six noise area categories specified, ranging from:

R1 - areas with negligible transportation, to
R6 - areas within predominantly industrial districts or with extremely dense transportation.

The Melbourne survey introduced intermediate ratings to provide more flexibility in the rating system. This modified system was given a trial in the 1983 Brisbane survey but was discarded after analysis indicated that its use did not result in any improvement in correlation with data in AS1055.

An assessor has the task of subjectively rating each survey site according to the above system, regardless of the actual zoning by the Local Authority. In general, sites in the surveys were selected in residential areas in either back or front yards at least 3 metres from residences. The number of sites selected for each survey ranged from six in the 1979 Sydney survey (for R2 area) to forty in the Melbourne survey. Relatively few sites were chosen in R1 and R6 areas. The average number of sites in each area category was approximately four, which is considerably less than a value of about ten preferred for reliable estimates of mean noise levels.

Variations in Noise Level

Most of the surveys were carried out over a single 24 hour period. The 1983 Brisbane survey was conducted for three days at the majority of the sites and the standard deviation of the mean Lₐₚ noise level at each site was less than 1 dB(A) in the standard time periods. The resulting 90 percent confidence interval for mean Lₐₚ levels was about ±2 dB(A). Since the method of assessing the area categories of sites is a subjective method, one would expect a larger variation among mean noise levels in an area category than among daily noise levels at a site. The value of the mean standard deviation over all the surveys was highest (3.4 dB(A)) during the Night period and lowest (2.4 dB(A)) during the Day period.

Regression Analysis

Regression analysis was carried out to determine how closely the survey data matched the AS1055 data. Values of the correlation coefficient were calculated for Lₐₚ values in the standard time periods and ranged from 0.37 to 0.95. Best agreement was noted for the Day period. Lines of regression for this period are shown in Fig. 1.

Values of the correlation coefficient have been shown to be significant at the 95 percent confidence level for the 1974 Brisbane, Melbourne and Adelaide surveys. For the 1983 Brisbane survey, significance was established for the Day and Morning periods only. Similar results were obtained for Lₐₚ values, but the values of correlation coefficient and significance level were slightly lower. Analysis of values of Lₐₚ and Lₚₚ also indicated significant correlation with tabled values of Day background level.

The magnitude of the standard error values indicated that, apart from data in the 1974 Brisbane survey, the data were rather dispersed and hence not highly reliable. Regression analysis of Lₐₚ and Lₚₚ values could show closer correlation with the AS1055 levels than do Lₐₚ values if the location of survey sites were standardised with respect to their distances from residences and positioning in allotments.

Mean Noise Levels

It is difficult to carry out any detailed comparison of noise levels among the various cities because surveys were conducted in different years. However, certain observations can be made about the data in Table 1 for R2 (i.e. quiet residential) areas.

<table>
<thead>
<tr>
<th>AS1055</th>
<th>BNE</th>
<th>MEL</th>
<th>SYD</th>
<th>ADE</th>
<th>BNE</th>
<th>SYD</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N) 40</td>
<td>36.8</td>
<td>37.8</td>
<td>-</td>
<td>42.9</td>
<td>41.8</td>
<td>-</td>
</tr>
<tr>
<td>(D) 45</td>
<td>38.4</td>
<td>40.5</td>
<td>36.3</td>
<td>43.1</td>
<td>44.0</td>
<td>39.8</td>
</tr>
<tr>
<td>(E) 40</td>
<td>35.0</td>
<td>37.5</td>
<td>39.6</td>
<td>40.1</td>
<td>42.6</td>
<td>36.4</td>
</tr>
<tr>
<td>(N) 35</td>
<td>31.0</td>
<td>31.7</td>
<td>33.3</td>
<td>35.3</td>
<td>37.8</td>
<td>32.6</td>
</tr>
</tbody>
</table>

Table 1. Mean Values of Lₐₚ - R2 Areas

(1) None of the Day values exceed the AS1055 limit of 45 dB(A).

(2) The relatively low Evening and Night values in the 1984 Sydney survey are believed to be the result of decreased industrial activity. The increase in the Day level has been shown to be significant at the 95 percent confidence level. The survey is being extended to include other area categories.

(3) In the comparatively recent Adelaide and Brisbane surveys, the Evening and Night values of AS1055 are exceeded.

(4) There has been a general increase in noise levels in Brisbane.
(5) The noise levels in the 1983 Brisbane and 1982 Adelaide surveys are generally similar. Analysis indicates that the differences in levels for each time period are not significant at the 95 percent confidence level. Similar results were obtained for R3 and R4 areas in Brisbane and Adelaide. Accordingly, the method of rating areas for surveys can be expected to yield equivalent mean values of $L_{eq}$ in the same time periods and area ratings for different cities.

Other Noise Descriptors

The 1974 Brisbane survey concluded that the widely accepted $L_{eq}$ criterion of 55 dB(A) for areas outside residences was equal to the mean value of $L_{eq}$ in R3 areas. The Adelaide and 1983 Brisbane surveys indicated that this criterion was exceeded in areas zoned R2 and above. A study of cumulative noise distributions for Adelaide and Brisbane indicated that the graphs were similar in shape to those reported in the USA [1]. In relation to the HUD criterion for daytime noise and that proposed by Challis [8] for night-time noise, the graphs of mean noise level all lie at worst in the "normally acceptable" region.

CONCLUSION

A description has been given of the Australian method of rating areas in order to establish noise criteria for planning. The $L_{eq}$ descriptor has been shown to be suitable for this purpose.

The author acknowledges the permission of the Director of Noise Abatement to publish this paper.

REFERENCES

ANOMALIES IN AUSTRALIAN ENVIRONMENTAL NOISE LEGISLATION

I. W. Eddington
DUIAE, Post Office, Darling Heights, Toowoomba, 4350, Australia.

INTRODUCTION

It is often claimed in Australia that noise pollution is now a major problem in our cities and that it is interfering with the quality of life of many people. But it is important to note that many claims are made in the presence of a legislative noise control management initiative which, being well into its second decade, is costly. It is time to ask whether the legislation in its present form is as effective as it might be in the fight to maintain environmental noise amenity.

Undesirable noise might well be an unavoidable disamenity of desirable goods and services chosen by the community. If this is so then "good" legislation may ever be at a loss to bring about improvement. However, "bad" legislation can most certainly hinder progress and might even allow deterioration of what otherwise might have been a maintainable status quo.

This paper tacitly assumes that the citizenry and its industries and institutions will, by desire or default, conduct their daily business mindful of the requirements of law. Furthermore, that noise legislation in Australia and reveals anomalies which, in certain circumstances, weaken the effectiveness of that legislation through the ambiguity and uncertainty they permit. The main body of the paper follows, and it outlines those anomalies as they emerge from description and analysis of the legislation. It also discusses implications of those anomalies. A conclusion is presented.

ANOMALIES IN NOISE LEGISLATION IN AUSTRALIA.

Introduction

Legislative control of community noise began in Australia in the first half of the 1970's. Since its beginning, noise control has remained a State responsibility, there being no national statute in the field. What little co-ordination there is, is exercised through the Environmental Noise Control Committee (ENCC) of the Australian Government. The Council itself is advisory and the ENCC is faced year in, year out, with scant financial resources relative to the responsibilities prescribed for it in its terms of reference. This lack of viable co-ordination constitutes the first anomaly: an anomaly of fragmentation.

Fragmentation

The implications of fragmentation are far reaching. First, because it permits a variety of measurements, assessment procedures and the like, it compromises the efficient and economical collection of national data, thereby removing comparability of data across states and, even more seriously, denying that vital data base necessary for evaluating the usefulness of standards, procedures and strategies, and the usefulness of the legislation itself.

Second, it permits a range of noise emission standards for products which are sold nationally. This indirectly adds costs to commodities as manufacturers must plan their products to meet a range of standards. Even if meeting the most stringent standard satisfies the rest there is still the administrative/research cost in searching and testing across the range of standards to ensure that this is so.

Third, the absence of a national standard works against freer international trade due to the smaller product and transport runs it requires. This absence of national standards also removes an effective noise abatement strategy for small partly industrialised nations like Australia. Such nations import a large proportion of their future noise environment when they purchase industrial machinery and equipment and transport equipment and it is difficult to push for stringent international noise emission standards when national ones are absent.

Fifth, it provides lawyers with a row of field days during litigation. When specified permitted levels for a particular type of occurrence differ between states, and where procedures conflict, skilful lawyers are able to obfuscate even the most explicit legislation. Sixth, it is wasteful with research funds, administrative time and legislative energy if personnel are paid to research which are the best methods from among a range of procedures and practices all of which are likely to change from time to time as perceived state needs dictate. Whereas the adoption of one practice and the subsequent eradication of its weaknesses by a combined co-ordinated effort from a number of places might soon lead to excellence, the present approach is akin to creating problems just to solve them.

Furthermore, when the fragmentation is carried on intrastate and responsibility for noise control is scattered over a number of Acts, a further implication emerges. When terms and definitions are not standardised across acts it is possible that many real complaints can not be effectively dealt with as they fall definitionaly between Acts. In these situations the aggrieved citizen is a Buridan's ass which, placed legislatively between two bales of hay, and, unable to choose between them, starves slowly to death under the full gaze of the lawmakers.

Seventh, fragmentation, by its very nature precludes economies of scale in the purchase of equipment, and the co-ordination of training and in the identification of common forward planning abatement techniques.

Noise control, and legislative strategy for noise control are complicated phenomena and there is no suggestion that the implications of fragmentation outlined above are an argument for complete centralisation; they are not. On the contrary, greater co-ordination of noise abatement activity might turn the anomaly of fragmentation into a strength, especially if it allows the individualised state approaches access to standardised procedures which are cost saving and which enhance future noise amenity.

Structure

A second anomaly is inherent in the manner in which Acts are structured to match executive ministerial decision making with the 20th century imperatives of method in science and technology. The majority of the acts are structured to contain a board or authority or so called commission between the minister and his scientific technical personnel. This group usually consists of personnel chosen for their position in industry (or commerce or government) and for their rare knowledge about interlinkages between various sectors of the economy. There are very few instances in which group members also possess expert knowledge in scientific and technical aspects of noise. Nevertheless it is very often the function of these groups to issue orders and licences.
and in fact to decide the more important actions taken under the Act. These boards are chosen by the relevant minister and have the power to co-opt scientific and technical staff in an advisory capacity. It is their unenviable duty to bring forward decisions which are fair and reasonable, which are in accordance with the Act, and which have, where possible, an empirical basis in science and technology.

Two implications follow. First, when standards, empirical measurement techniques, and permitted levels are not specified and binding within the Act, it becomes again very easy for skilful lawyers to play on the issue of one persons rights against another persons rights rather than on each of the persons rights against a law based upon empirical fact and tested procedure. In time, decisions made under this system can often become contradictory (as also do court decisions on appeals) and in this manner the law can be rendered less effective in the presence of a measurable increasing noise disamenity. Second, the method of the interposed board makes costly scientific training and expert knowledge the second tiddle, and clouds the contention that an Act which sets binding (on the judge as well) performance standards written in terms of empirically tested categories of activity and emitted noise, might more effectively ameliorate the slow upward movement of noise disamenity.

Time

Although a number of types of noise complaint can be dealt with quickly and efficiently under the Acts, there are circumstances in which environmental noise disamenity is able to continue unabated owing to the frequency and regularity with which authorities and boards are able to meet and because of the time passage allowed for appeal and/or the hearing of appeal. Of course these permitted times are an acceptable and reasonable convention in all enlightened legal systems and it is only when orders and the like are inoperative pending the outcome of the appeal that they can be legally used to allow the continuation of otherwise excessive noise, sometimes for as long as is necessary to complete the job. Construction sites are an example.

The anomaly of time referred to here is very closely associated with the absence of procedural policy in key areas of noise making activity and also with the absence of specified levels and standards upon which scientific, technical and administrative staff might operate to obtain a clear cut decision. In the present state of the art the anomaly of time is not an easy one to remove. The task of constructing those largely absent standards and procedures and empirical databases necessary for obtaining accurate objective answers about a wide range of complex noise situations is a task of huge proportions made difficult at the outset by the lack of data resulting from the fragmentation discussed earlier.

Standards

The extent to which standards and procedures are called up varies considerably between Acts but generally the legislation can be said to be disadvantaged initially from an absence of relevant standards upon which to call in the first place, and again from an unwillingness to use existing standards in a decisive way and finally from ambiguities within existing standards themselves; Australian Standard 1065 being an example, where it stops just short of defining noise $\text{SdB(A)}$ above background as nuisance noise.

Ambiguity, together with limited use of standards has a number of implications. First various states adopt their own methodologies thereby increasing fragmentation and uncertainty and administrative cost. Second the existing standards remain less useful as they are not subject to that gradual refinement brought about by continual use. Thirdly, ambiguity within the standards leads to greater subjectivity in the interpretation of noise situations. While one may well argue that subjectivity is desirable in resolving noise problems it also becomes the Achilles heel in litigation. The absence of standards and other objective procedures in many vital areas of noise control removes that objective basis with which to peg down and hold down, ambient noise levels.

CONCLUSION

Four anomalies inherent in Australian environmental noise control legislation were separately examined. Each of these anomalies was shown to be potentially damaging to the effectiveness of the legislation - especially if the aim of that legislation is to hold steady, or even reduce noise levels, rather than simply to control them. But, de facto, the anomalies discussed are really inextricably interwoven and the manner in which they sometimes combine to undermine effectiveness is another study in itself.

An interesting question emerges from the notion that to remove the anomalies would improve the legislation's chance of success. A tighter legislative package might impose costs at a rate greater than society can pay and the abrogation of responsibility that this might bring could result in a legislation less effective than it now is.

Legislative noise management is really a modern experiment and there must be many working in the field who wonder whether or not it is a subject more complex and difficult than was at first imagined.
THE STATE OF NOISE CONTROL IN ARGENTINA

G. L. Fuchs
Av. Colon 845, (5000) Cordoba, Argentina

Noise control is one of the branches of Acoustics of greatest priority in Argentina. Nevertheless it is not fully developed in human resources and technological backing.

We have carried out a survey on the academic development, research in progress and availability of acoustical materials in the main branches of Acoustics. The results of our survey are shown in Table 1. For comparison we have included the courses offered at 92 universities of North America (Ref 1) in which noise control emerges as the highest priority. Curiously enough Architectural Acoustics ranks second in Argentina and last but one in North America (Ref 2).

We have achieved considerable progress through legislation, especially on Safety and Health at national level. A law was passed (N 19587 regulated by decree 551/79) based on our report published a few years ago. Trade Unions and the Ministry of Labor put considerable pressure on industry resulting in a lot of practical progress. The author proposed municipal legislation and got it approved as Ord. 497/70, which pioneered in Latin America, as far as Mexico.

Vehicle noise limits and neighborhood noises have been set based on international trends: levels in dB(A), scarce peaks L10 and frequent sound L90, as the climate.

The author introduced as our Code of Practice (Cordoba, industrial center, pop. 1 million) isolation limits between units of vertical and horizontal property, avoiding control difficulties we met in indexes by model acceptable partitioning and floating floors. Local inexpensive elastic materials such as polystyrene and surfacing covered by an impervious film on which screwed allow various floor finishes. These floating floors of high efficiency add less than 1% to the total construction cost in our country. Very poor response from both builders and owners may be ascribed to lack of knowledge and routine. But I must admit I have heard footsteps and impacts both in North American and European buildings!

About absorption there is a widespread belief that an acoustical ceiling will perform the triple miracle of insulating step from upsetters; muffles neighbor's noises and isolate street traffic noise. The latter exceeds L90 most of the day among us with bedrooms opening onto the street. These "acoustic ceilings" are commercially advertised and sold as a Panacea. But it is fair to admit a sustained increase in the number of adequately isolated buildings.

Less successful has been the treatment of mechanical installations, the lack of piping and ducts isolation from structure. Air conditioning, water and sanitary contractors resist consulting acousticians until it is too late to correct major faults.

Another item almost never included in the noise control checklist are elevators, especially the machinery housing, from which impacts and vibrations propagate in the form of bending waves later re-emitted as sound at considerable distances. The annoyance of start-stop noises of elevators are only second to footsteps as sources of nuisance mainly in multistory buildings.

Our type of construction has acoustical advantages and disadvantages. Generalized use of brick masonry and concrete provides high insulation but creates reverberant enclosures in which the clattering is hard and flooring is hard wood or ceramic tile. We have measured monotonously sized, furnished enclosures in which reverberation times are well above the usual 0.5 seconds. Flutter echoes are quite common and more noticeable in large rooms and corridors.

It is hard to convince most architects and contractors to include acoustical treatment in the building generally rather go in for fancy bathroom and kitchen fixtures, tiling, mirrored walls and windows. The latter is one of the major problems of building acoustics in our country: "transparent" façades. Designers are set on letting occupants "enjoy the view" even on many a time too far from poetic. Newspaper advertisements read "all bedrooms looking on to the street" as a sales feature.

A short anecdote about a fully glazed façade: during a meeting at the 20th Floor of a São Paulo, Brasil skyscraper, communication was quite difficult. The subject: Noise Control. My talk was on priorities in building insulation, so I had a red hot example at hand when ranking façades as priority ONES: it is not easy to talk designers into blind, highly isolating façades, but there are reasonable alternatives. An Anita Lawrence pointed out in her 9th IGA paper (Ref 4) open windows rarely isolate more than 10%. And in sub-tropical countries it is inevitable to have people opening windows to breathe in the scented Spring air. Double glazing implies air conditioning and fixed windows, different crystal thickness, internal absorption, etc and costs soar up.

Some have succeeded in keeping façades for garages and halls at ground level and service or meeting rooms requiring privacy, etc, letting the "scenic" and "refreshing" windoows opening on to internal "patios". It is interesting to note that Spanish Colonial architecture in our countries featured small openings at the front while life was enjoyed at one or more internal patios.

Among us roofing is massive enough to isolate noise from overflights and far away sources (highways and industry). Teaching areas opened on to streets with tall windows, but the traffic was not annoying. Rather the reverberant interiors deteriorated intelligibility. Modern classrooms and auditoria are kept away from the noisy areas and absorption is widely used especially in ceilings.

Music rooms and auditoria (which we discuss further in another paper by Oomen and Fuchs) are nowadays multipurpose and acoustically designed.
We have some fine classical theatres and Opera houses designed in pre-acoustics times. But designers followed the European trend (nonsense voice, rococo ornamentation plenty of carpeting and curtains) and they sound well except when the stage is too big. (see paper above mentioned, J. Osuna & Fuchs)

Radio broadcasting stations and T.V. sets almost without exception are acoustically designed in teamwork with architects, engineers and contractors.

Industrial Noise

Industry started in Argentina about a century ago without the least regard for acoustics. You can still find highly reverberant sheds with R.T.S. of 5 or more seconds housing noisy machinery generating 95 to 110 dBA as measured by us in the early attempts at correcting industrial noise by acoustical means. As expressed at the beginning of this paper the 90 dBA/AHR limit was anxiously requested by industrialists under legal pressure. Lack of experts lead to clumsy attempts at lining walls with the then called "thermo-acoustic" pads, but not less than 15 years ago, functional absorbers and later partial and total enclosures were designed properly. Vibration isolation helped much by limiting propagation.

To restrict such an extensive subject we shall limit our discussion to control of aerodynamic and impact sources which are the biggest offenders in our plants. I would like to mention among many others an industrial plant where half a dozen outlets from steam generator turbines produced a combined level of 140 dB. We combined all outlets into a silence chimney and could reduce the noise which reached a 2 km distant village at a level of almost 90 dBA depending on the wind, to background level of the village.

Among impacts we found drop forge hammers the most difficult to control. A car factory's forging plant was improved by large isolated foundations plus absorbing partial enclosures plus compulsory personal protection. Rest periods in a treated room where the level allowed recovery without creating a shock (thermal and acoustical) when resuming work helped a lot in hearing conservation of the personnel.

Summarizing, after many years of changing hearing conservation criteria and indices, we have simplified to a minimum the criteria according to 4 sensitivity steps for all types of activity (Table 2).

(1) A.S.A. Committe. . See Ref. 1. Ext.
(2) A.S.A. Committee. . See Ref. 1. Ext.

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>FUNCTION</td>
</tr>
<tr>
<td>INDUSTRIAL PLANTS</td>
</tr>
<tr>
<td>VEHICLES (TRANSPORT)</td>
</tr>
<tr>
<td>URBAN SERVICES</td>
</tr>
<tr>
<td>SOCIAL ACTIVITIES</td>
</tr>
<tr>
<td>CREATIVE WORK THINKING</td>
</tr>
<tr>
<td>STUDDING</td>
</tr>
</tbody>
</table>

TABLE 1

<table>
<thead>
<tr>
<th>UNITS</th>
<th>TEACHING</th>
<th>RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANG.</td>
<td>ACU.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Arch. AAS.</td>
<td>Insulation (NR)</td>
</tr>
<tr>
<td>(ASA 30)</td>
<td>Absorption Auditories</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>Noise Control (ASA 40)</td>
<td>Community Industrial Buildings</td>
</tr>
<tr>
<td>100</td>
<td>Electroac. (ASA 60)</td>
<td>R. &amp; D.</td>
</tr>
<tr>
<td>90</td>
<td>Psychac. (ASA 60)</td>
<td>Speech Analyses</td>
</tr>
<tr>
<td>80</td>
<td>Oceanography (ASA 60)</td>
<td>Propagation</td>
</tr>
<tr>
<td>75</td>
<td>Physiological Acoustics (ASA 28)</td>
<td>Cochlear stimulation</td>
</tr>
<tr>
<td>60</td>
<td>Signal Processing (ASA 60)</td>
<td>Audio/Video conv.</td>
</tr>
<tr>
<td>80</td>
<td>Vibration Shock (ASA 40)</td>
<td>Vibration anal.</td>
</tr>
</tbody>
</table>

Notes: Illustrations will be shown as slides during presentation.
A REVISION OF THE ERCB
NOISE CONTROL GUIDELINE

C. Andrew and E.H. Bolstad

City of Calgary, Transportation Department
P.O. Box 2100, Station "W"
Calgary, Alberta, T2P 2M5

Bolstad Engineering Associates Ltd.
9249 - 48 Street
Edmonton, Alberta T6B 2R9

The Energy Resources Conservation Board (ERCB) was established by the Government of Alberta, Canada to ensure the safe and efficient development of energy resources in Alberta. In its authority to regulate the operation of energy facilities, the ERCB reviews all major developments to ensure that standards for safety and pollution control are met.

In the absence of Provincial industrial noise standards the ERCB established maximum permissible noise levels applicable to energy resource industry operations such as compressor stations, drilling and production of oil and gas installations in rural areas which previously enjoyed quiet living environments. Noise levels ranging from 55 to 70 dBA have been experienced in areas accustomed to levels in the 35 to 50 dBA range. In view of this disparity the ERCB directed that 1080-2 be updated and revised to take into account concerns voiced by the affected public and energy industry representatives including acoustical specialists involved in noise impact assessment and control. A committee of knowledgeable representatives of various concerned organizations was formed. The objective of the Noise Control Committee (NCC) is to revise the current guideline 1080-2 and include recent developments in acoustic technology and noise control standards.

PURPOSE

The main purpose of this report is to outline the procedure or methodology followed in the formulation of the revised noise control guideline and to highlight the recommendations resulting from this process. The guideline is composed of a number of sections and each is discussed to content and context. It must be noted that this report outlines a draft of the revised guideline which will serve as an "interim directive". Suggestions for amendments are therefore encouraged and will be closely considered before it becomes a 'regulation' in its final form.

NOISE CONTROL GUIDELINE REPORT

Preamble

In order to provide an introduction to the guideline, a preamble is necessary so that the user can be clear as to the reasoning and philosophy behind the guideline. Some of the main points stressed in the preamble are that the intent is to allow flexibility in dealing with energy facility noise and designs of noise control methods; that the consideration for noise control in the earliest stages possible is the most desirable goal to achieve; that the guideline addresses the impacts of noise on people and that investigations for compliance will be conducted for the most part, on a complaint basis.

Discussion

The Committee held a number of meetings to discuss the various issues to be considered. Once an overall philosophy was established, a number of sub-committees were formed to address specific questions. Included in these discussions were such issues as acceptability criteria, limitations of the guidelines, enforcement and jurisdiction, facility related traffic noise, existing versus future facilities and "creeping ambient", instrumentation and measurement conditions.

Acceptability criteria and the resultant Basic Noise Levels (BNL) were developed on the premise that land use (categories) and residential density pre-determines the ambient levels of an area and the degree of intrusiveness which would be acceptable. Adjustments to these levels are allowable because of variations in the operation of a facility as well as variations in the sensitivity of the receiver. Therefore, depending on the ambient level, season, temporary versus permanent operation, presence of tonality and impulse/impact noise, etc., a number of adjustments could be applicable but with a maximum allowable.

Traffic noise related to a facility was found to be a necessity and the intent of an adjustment was to allow some limited noise but when combined with facility noise, the overall level was to be controlled by industry.

Enforcement and jurisdiction was a concern from the standpoint of the ERCB having jurisdiction over traffic on a roadway and from a facility which is typically under local municipal jurisdiction. It was agreed that local municipalities were not equipped to handle this problem and that as long as the ERCB maintained ongoing liaison with local authorities, continuity would be maintained thus minimizing conflicts in jurisdiction. The concept of "creeping ambient" and industry development was discussed. It was determined that the guidelines were intended to control "creeping ambient" to a certain extent but without prohibiting the development of new industry entirely. Other factors such as economic development and the local economy would most likely outweigh concerns with noise impacts in most cases.

Design And Compliance Noise Levels

This section outlines the Maximum Allowable Noise Levels (MANL) for both design and compliance applications. The MANL are to be used in the design of a facility or in the planning of an operation and shall be calculated 15 metres from the nearest or worst impacted dwelling unit. This level shall consist of the Basic Noise Level (Table 1) and the Class A and Class B Adjustment Design Values (Table 2).

Compliance Noise Monitoring shall be performed when a complaint is received and Compliance is determined by comparing these levels to Basic Noise levels. The Basic Noise Level has been set in an allowance of 5 dBA for existing industry.
Table 1: Basic Noise Levels (dBA Leq)

The Basic Noise Levels presented here are intended to be representative ambient noise levels due to local activities characteristic of the area plus an allowance for industrial presence.

<table>
<thead>
<tr>
<th>Transportation Proximity</th>
<th>Daytime</th>
<th>Nighttime</th>
<th>Daytime</th>
<th>Nighttime</th>
<th>Daytime</th>
<th>Nighttime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1</td>
<td>10-30</td>
<td>40</td>
<td>25</td>
<td>35</td>
<td>20</td>
<td>25</td>
</tr>
<tr>
<td>Category 2</td>
<td>35</td>
<td>50</td>
<td>25</td>
<td>50</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td>Category 3</td>
<td>50</td>
<td>60</td>
<td>40</td>
<td>50</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Category 4</td>
<td>60</td>
<td>70</td>
<td>50</td>
<td>60</td>
<td>50</td>
<td>70</td>
</tr>
</tbody>
</table>

Daytime Level: A-weighted continuous energy equivalent sound level from 0700 to 2300 hr.
Nighttime Level: A-weighted continuous energy equivalent sound level from 2300 to 0700 hr.

Table 2: Adjustment Values

<table>
<thead>
<tr>
<th>Class</th>
<th>Adjustment(a)</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Seasonal Adjustment (Nov 1-Mar 31)</td>
<td>5</td>
</tr>
<tr>
<td>A2</td>
<td>Heavy Truck Adjustment (from Table 3)</td>
<td>5</td>
</tr>
<tr>
<td>A3</td>
<td>Absence of Total and Impulse/impact Components Adjustment</td>
<td>5</td>
</tr>
<tr>
<td>A4</td>
<td>Ambient Monitoring Adjustment (from Figure 1)</td>
<td>5</td>
</tr>
<tr>
<td>Class B Adjustment = Sum of A1, A2, A3, and A4 (if applicable), but not to exceed a maximum of 10 dBA Leq.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Temporary Activity Adjustment for total activity complete within:
- 1 day: 5
- 1 week: 10
- 2 weeks: 15
- 3 months: 20
- greater than 1 month: 25

Class B Adjustment = one only of B1, B2, B3, or B4

1. Applicable only upon submission of technical verification.
2. Applicable only within 1 km radius of the facility.
3. Usable only when audible characteristics of a permanent facility are absent of both common and impulsive or impact components. Absence of Total Components shall be when the measured 1/3 octave band, slow response sound pressure levels of any one or two adjacent bands (within ± 1 oct) are 80 dB or higher than the lower limit of the band.
4. Absence of Impulse/Impact Components shall be when the difference between the A-weighted Impulse response sound level measurement and the A-weighted slow response sound level measurement is 15 dBA or less.
5. Usable only when Basic Noise Levels (Table 1) may not be representative of actual ambient noise environment. Applicable only when ambient noise monitoring survey has been conducted prior to application approval. Ambient noise monitoring shall consist of 24 hour continuous monitoring survey, with measured ambient noise levels presented for the daytime and nighttime periods, conducted 15 metres from the nearest or worst impacted dwelling unit.
6. Cumulative duration of periods of noisier operation during temporary activity.

Summary and Conclusion

The overall intent of the revision of 1080-2 is to improve the existing and future noise environments of residential areas in the vicinity of energy related facilities in Alberta. This report has touched on a number of key issues discussed throughout the process of developing this revision. The guidelines described herein is a draft proposal and has yet to be approved by the ERCB. Therefore, although the Committee covered the main aspects related to the scope of the revision, comments are invited.

Acknowledgements

I wish to thank a number of participants in the Noise Control Committee for their assistance in the preparation of this paper. Mr. H. Abbink of ERCB, Mr. Leslie Frank of HPC Acoustical Consultants and Dr. B. Balachander of Western Research.

Table 3: Heavy Truck Adjustment A2

<table>
<thead>
<tr>
<th>Number of Truck Passes as a Distance Less Than 30 Metres</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Noise Level (Table 1)</td>
</tr>
<tr>
<td>Nighttime (Leq 9 hrs.)</td>
</tr>
<tr>
<td>Daytime (Leq 15 hrs.)</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>45</td>
</tr>
<tr>
<td>Number of Truck Passes as a Distance More Than 30 Metres但 Less Than 100 Metres</td>
</tr>
<tr>
<td>Number of Truck Passes (Table 1)</td>
</tr>
<tr>
<td>Nighttime (Leq 9 hrs.)</td>
</tr>
<tr>
<td>Daytime (Leq 15 hrs.)</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>40</td>
</tr>
<tr>
<td>45</td>
</tr>
</tbody>
</table>

1. For number of truck passess falling between values use next highest passess value.

Heavy Truck - vehicle related to facility development or operation having three or more axles.
LA COMMUNE ET LA LUTTE CONTRE LE BRUIT

André CROS
Société Acoustique et Conseil - 36 rue George Sand - 95500 Rueil Malmaison - France

INTRODUCTION

C'est en particulier pour les communes de plus de 20 000 habitants, qui sont au nombrissement de 12 000 en France, dans lesquelles résident 80% des Français, que le Ministère Français de l'Environnement a demandé en 1985 à la Société Acoustique et Conseil, dont André CROS est le créateur et le Directeur, de réaliser une étude dont le but est de donner aux élus et aux Services Municipaux de ces communes un outil méthodologique complet et concret pour aborder et traiter globalement la nuisance brute, par la mise au point d'un Plan Municipal de Lutte Contre le Bruit.

L'IMPORTANCE DU PROBLÈME BRUIT DANS LA COMMUNE

Quelques chiffres permettent de situer l'importance et la progression du problème brut :-
- Le Ministère de l'Environnement a enregistré entre 1983 et 1984 une progression de 42% des plaintes qui lui ont été adressées (1 232 en 1983; 1 745 en 1984).
- De leur côté, les Préfectures dans les Départements ont enregistré 9 100 plaintes en 1983 et 10 400 en 1984.
- En ce qui concerne les Villes, le nombre de plaintes dans les 400 communes françaises de plus de 20 000 habitants, a pu être estimé à 20 000 en 1985.
- A titre d'exemple, la Mairie de Grenoble, Ville très connue, a reçu et traité en 1985 1 200 plaintes concernant le Bruit.

LA TYPOLOGIE DES BRUITS DANS LA COMMUNE

Au-delà de ces chiffres, les élus et fonctionnaires Municipaux savent bien dans leur activité communale combien le Bruit est un danger durement ressenti sous tous ses aspects :
- Bruits des transports et de circulation : deux-roues, camions, voitures, avions, chemins de fer, autoroutes...
- Bruits de la vie communale et collective :
  # Vis à vis de l’extérieur : fêtes foraines, fêtes, festivals, manifestations sportives, discothèques, marchés, rassemblements d'ordres, etc...
  # Vis à vis de l'intérieur : bruits dans les crèches, les gymnases, les restaurants et préaux scolaires, les salles polyvalentes, les bureaux d'accueil, etc...

Il est indispensable que la commune se fixe ensuite, à échéance de la durée de son plan, des objectifs en terme d'amélioration pour les différents types de bruits générés. Ces objectifs seront, pour être réalisés, plus qualitatifs que quantitatifs.
TROISIÈME PHASE - LE PLAN D’ACTION MUNICIPAL
DE LUTTE CONTRE LE BRUIT :

La phase essentielle consiste alors à
fixer un plan d’actions susceptibles de
permettre à la Ville d’atteindre les
objectifs qu’elle s’est assignés. Ces
actions seront tidées aux 8 principes déjà
énoncés et seront regroupées dans les 4
chapitres suivants :

- Actions de sensibilisation et d’information
- Actions de prévention, contrôle, et
réglementation
- Actions techniques (urbanisme, études et
travaux acoustiques, etc...)
- Actions de formation technique.

Ces actions feront l’objet :
- d’une définition (cahier des charges)
- d’une planification
- d’un chiffre (éventuellement)

Le Plan d’action comportera un chapitre
spécial consacré aux moyens mis en place
pour l’exécuter :

- moyens en dépenses de fonctionnement
- moyens humains pour l’instruction des
plaintes, la prévention, le contrôle, la
réglementation, la formation du personnel
communal, les études techniques et
acoustiques, etc...

Tel est le schéma de mise au point d’un
vrai plan de lutte à moyen terme contre
le Bruit dans la commune.

Reste bien entendu l’essentiel, à savoir
l’exécution, le suivi et le contrôle de ce
plan sur toute sa durée, en associant les
compétences et les bonnes volontés internes
à la commune, avec un Conseil Extérieur qui
assiste et guide la commune, comme le fait
eux jour en France, une Société comme
Acoustique et Conseil.
DRAFT NATIONAL ENVIRONMENTAL NOISE CODE

D.A. Benwell

Non-Ionizing Radiation Section, Bureau of Radiation Protection and Medical Devices, National Health and Welfare, Room 228, Environmental Health Centre, Tunney's Pasture, Ottawa, Ontario K1A 0L2

INTRODUCTION

This paper describes the Canadian draft National Environmental Noise Code (NENC or Code) which is being prepared to provide information to municipalities, consultants, provincial planners, industries, designers and to legislators at all levels of government. It is being prepared by the Working Group on Environmental Noise for the Federal/Provincial Advisory Committee on Environmental and Occupational Health.

The draft Code presents methods for the assessment and exposure to and control of environmental noise. The purpose is to provide a common basis for this work across Canada, while at the same time providing options for flexibility of choice to account for specific needs.

Although a similar code at this time, the Organization for Economic Co-operation and Development (OECD) has been active in providing guidance to strengthen Noise Abatement Policies in the area of environmental noise (1). Canada has no present National Environmental Noise program, unlike countries such as France, Germany, Switzerland, and Australia (1).

NOISE SOURCES

Transport is by far the major source of environmental noise, with road traffic the chief offender (2). Aircraft noise comes next in terms of population exposure. This is particularly high in North America where the proportion of population exposed is four times higher than in Europe or Japan. Railway noise and noise from fixed sources such as industrial establishments generally affect a more limited portion of the population (3).

Noise levels may vary considerably from region to region, from town to town and even from one district to another within the same town. Factors such as population density can play a very significant role. Generally speaking, noise exposure levels are higher in urban areas, which are more densely populated and where traffic density is therefore higher.

Four distinct kinds of environmental noise sources can be identified. The first are stationary sources, for which the specific location can sometimes affect the significance of the noise on people. The primary example of this type of noise source is air conditioning equipment, although various kinds of industrial equipment can also be of concern in residential areas and noise sensitive land uses. The remaining three kinds of noise all stem from mobile sources: transportation noise, power noise from tools, and finally, people and animal noises. The most prevalent transportation sources are motor vehicles; including cars, trucks, buses and motorcycles which can be heard almost everywhere. Other transportation sources, which affect fewer people but may affect them more severely, are airport operations, railroad operations, and in some locations ships, motor boats, or recreational vehicles. Power tools are considered a source of community noise to the extent that they affect people other than their operators. Power lawn mowers are an obvious example. The final kind of noise which affects people in residential communities is noise caused directly by men and animal activities; children playing, outdoor parties, dogs barking, etc.

It is fair to note, especially because of the variety of sources of community noise, that their control needs to be flexible.

DESCRIPTION OF CODE

The draft National Environmental Noise Code is basically divided into 8 sections (see Table 1). The first two sections describe the purpose and organization of the document and the present government jurisdictional arrangements into which it fits.

<table>
<thead>
<tr>
<th>Table 1. Outline of Draft National Environmental Noise Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Introduction</td>
</tr>
<tr>
<td>2.0 Present Government Jurisdiction</td>
</tr>
<tr>
<td>3.0 Definitions and Interpretations</td>
</tr>
<tr>
<td>4.0 Instrument Specifications. Measurement Methods and Models</td>
</tr>
<tr>
<td>5.0 Noise Control Program - Options and Alternatives</td>
</tr>
<tr>
<td>6.0 Noise Control Program Implementation</td>
</tr>
<tr>
<td>7.0 Model Noise Control Legislation</td>
</tr>
<tr>
<td>8.0 Noise Reduction Techniques</td>
</tr>
<tr>
<td>9.0 Bibliography</td>
</tr>
</tbody>
</table>

Since environmental noise is associated with numerous activities in the community such as industry, commerce, transportation and residential activities, all of which are vital for the well-being of communities, control of noise may be conducted at any level of government, although it can be more effective and efficient if all 3 levels are involved. A summary table of the distinct responsibilities regarding noise control and abatement of each of the three levels of government in Canada is given in Table 2.

Sections 3 and 4 of the Code provide a common basis for language and for instrument specifications and measurement methods. Specifically, Section 3 gives definitions common throughout the document to assist those using the Code to develop their own regulations or by-laws or to use as guidelines to control environmental noise. Section 4 sets out the minimum specifications for equipment used for the measurement of sound and vibration and draws on Canadian and International Standards as appropriate. It also lists the various measurement procedures to be used in connection with the limits or standards set out in the Code.

The main action portion of the Code is contained in Sections 5, 6 and 7. Section 5 provides a summary of the options and alternatives available to the various levels of government to control environmental noise. These include preparations for a comprehensive noise control program, environmental noise legislation options (legislation of new sources, receivers, and abatement of existing noise problems), and municipal control of environmental noise. Table 3 illustrates the type of action plan being developed for environmental noise sources.

Section 6 concentrates on the implementation aspects of noise control, especially when dealing with land use planning. It is also divided into new sources, new receivers, and abatement of existing noise problems.

Section 7 presents model noise control legislation. Such legislation consists of 3 basic types of standard: (i) sound emission standards for specific products or sources, (ii) property line standards; and (iii) point of reception standards. Sample model noise control by-laws are provided that may be enacted by individual municipalities under
### Table 2: Jurisdictional Responsibilities and Activities in Environmental Noise

<table>
<thead>
<tr>
<th>ENERGY HYDROCARBON PIPELINES</th>
<th>AIRCRAFT AND AIRPORTS</th>
<th>RAILWAYS</th>
<th>NAVIGABLE WATERWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FEDERAL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Primary</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Activities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- noise emission standards and regulations of manufactured products, equipment...etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- noise emission standards and regulations of facilities operated under federal charter as well as inter-provincial transportation vehicles</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- national building code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- federally assisted housing (CMHC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- retrofit programs and home improvement programs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- environmental assessment of projects subject to federal legislation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- other activities under federal legislation (Environment, Health &amp; Welfare, Transport...etc.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PROVINCIAL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Activities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- acts, regulations, policies, codes, guidelines, model by-laws...etc. that deal with planning, building, environment, natural resources, highway traffic, health, etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- assist municipalities in formulation, adoption and implementation of municipal by-laws</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- retrofit programs and home improvement programs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MUNICIPAL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Activities</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- municipal noise control by-laws</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- municipal land use plans and zoning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- traffic management</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- retrofit and home improvement under local or provincial, or federal improvement programs</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The final two sections describe the various noise reduction techniques available and give a bibliography of related texts and noise standards.

A 5 year review of the document will be recommended with updates at that time if required.

### Conclusions

The National Environmental Noise Code is thus a comprehensive document encompassing all aspects of environmental noise control and providing a much needed national basis for this activity. It is designed to allow flexibility of options to the user, and may be modified with the development of national environmental noise measurement standards, changes in technology or as other future developments occur. It may also be considered for publication as a National Standard of Canada.

### References

NOISE MEASUREMENTS AND ATTITUdINAL SURVEYS OF NIGERIAN CITIES

C.A. Bakare
University of Ibadan, Ibadan, Oyo State, Nigeria

INTRODUCTION

Noise pollution has become an increasingly important environmental problem of modern living. In several advanced countries of the world, the impact of noise has become a matter of great concern to the various governments and hearing health professionals. It is a matter of regret however, that very little or no attention has been given to the potential hazards of noise in developing countries.

Prior to the discovery of petroleum some two decades ago, Nigerian cities like most other African cities were relatively noise free. The trend at present however suggests a movement, in the so called developing nations, towards acquiring several conveniences of modern civilization. Such a move has brought in its wake the introduction of powerful industrial machineries, automobiles generating loud noises, motorists blaring their horns with reckless abandon, record and stereophonic stores competing for attention by playing the latest disco records at the optimum levels, all this going on side by side with noises generated by construction machineries during the daytime and at night, due to constant power failure, very powerful electric generators, noise generated by landlords to usher in the noise pollution night shift.

Since most people in Nigeria would not recognize noise as an insidious pollutant or attribute it serious physiological impacts, noise pollution studies have not been accorded a high priority. From the preliminary noise survey conducted at some of Nigeria's city intersections, it was found that noise pollution abounds everywhere. As no other studies of this nature have been reported in Nigeria, it was decided to extend the noise survey to cover four major cities in Nigeria. These are Lagos, the capital city of Nigeria, Ibadan, Enugu and Kaduna, with a combined population of over 8 million. Since there is a lack of consistent town planning, the cities are thoroughly mixed with so clearly defined industrial, commercial or residential areas.

PURPOSE OF THE PRESENT STUDY

The study aimed at obtaining the following objectives: (a) to assess people's reaction to environmental noise (b) to carry out noise measurements and identify their sources (c) to compare the noise conditions in the four Nigerian cities (d) to create an awareness as to the health hazards of noise pollution (e) to provide a background for the promulgation of noise control legislation in Nigeria.

METHODOLOGY

Attitudinal Survey

This survey which involved the completion of a 4 questionnaires is similar to that conducted by Bragon and Stathis' but modified to make it relevant to Nigeria. The three pronged questionnaire attempted to obtain information on the residential environment, individual's reaction to noise and demographics. To avoid potential biasing effects, the survey was not introduced as a noise survey and the questions relating to noise were deliberately omitted in the first section of the questionnaire.

A total of 800 questionnaires were completed over a period of six months at 16 designated stations, with 50 completed questionnaire per station. The sample size consisted of 61% males and 39% females aged 20 to 50 years. A stratified sampling technique was used so that all social classes were represented in reasonable numbers (Table 1).

Noise Survey

Although the most desirable method for measuring environmental noise is to record all intrusive sounds continuously for a period of 24 hours, the cost of such a procedure is prohibitive. Hence, the environmental sound sampling method of recording short duration samples have been found to be a viable alternative.

Since no systematic traffic volume record is available in any part of the country, the criteria for establishing the sampling periods were based on two known factors, i.e. the general time periods of the day (morning, afternoon, evening and night) and periods of high, medium and low mobility. Population mobility in Nigeria coincides with movement to and from work between 7 a.m. and 9 a.m.; 3 p.m. to 5 p.m. (high mobility) and 8 p.m. to 10 p.m. (low mobility). Mobility is reduced to almost nil during the night in the cities because of rising crime wave and police/ army road blocks.

Noise samples were taken at each of the 16 stations (i.e. 4 stations per city) on 4 different days, Monday and Wednesday (working days for all) Saturday (workforce day for civil servants but not for commercial houses and private enterprises) and Sunday. Measurements were taken during the 3 sampling periods (7 a.m. to 9 a.m., 12 to 2 p.m. and 8 p.m. to 10 p.m.) for each of the four days. A total of 12 visits were made to each station and at every visit, a noise sample was obtained for a period of 5 consecutive minutes. Fifteen sound level values were read directly from the meter, every 20 seconds, from which the arithmetic average was determined and subsequently the mean sound level was computed.

Acoustic Instrumentation

Noise measurements at the various locations were obtained by a portable sound level meter (Tracer RA-305) and was periodically checked with an acoustic calibrator. All measurements were obtained in good weather at street level with microphone placed 1.5m above the ground on sidewalks and away from reflecting walls where feasible.

RESULTS AND DISCUSSION

Analysis of Attitudinal Surveys

Table II presents the people's judgement of the degree of noise present in their environment. Only 6.5% of all respondents judged their environment very quiet, 11.2% considered it fairly quiet, 22.5% slightly noisy, 40.25% noisy and 19.5% very noisy. A total of 82.25% of all surveyed in the four cities were affected by noise.

In judging the degree of the problems affecting their environment, crime was ranked first, this was closely followed by noise problem, littering, poor housing and the least problem was air pollution (Table III).

More than 60% of those surveyed found vehicle traffic to be the most annoying source of noise followed by noise from the record stores (Table IV).

Amongst those surveyed, over 55% indicated that
they were more annoyed by noise when they were inside the house than by noise outside - 42% expressed annoyance at noise from the electric generators particularly at night.

Over 80% of those interviewed felt that noise affected them adversely but only 3% were taking precautionary measures to protect themselves from excessive noise. Amongst the specific disorders reported by those exposed to noise were headaches, dizziness, nervousness and lack of concentration. Almost all the people surveyed would support legislation to curb noise pollution.

Analysis of the Acoustic Survey

The mean noise level obtained from the four Nigerian cities suggest that the cities particularly Lagos, ranked amongst the noisiest in the world. The mean noise level in Lagos was 90.0 dBA, Ibadan 81.3 dBA, Kaduna 78.6 dBA and Enugu 74.0 dBA (Table V). Based on the data from several studies and the "Levels Document" suggested by the EPA, Lagos has exceeded the 70 dBA weighted maximum, safe, yearly sound level with respect to noise-induced hearing loss by as much as 20.0 dBA and the least noisy city (Enugu) by 4.0 dBA (Figs. I and II).

CONCLUSIONS

Although the data collected in this study have limited scope due to financial constraints and therefore liable to some errors, it is abundantly clear however that the Nigerian cities surveyed are noisy and that the Nigerian people are exposed to noise sufficient enough to impair health.

A lot of people agree that noise is a major environmental problem but very few are aware of its health hazards. Due to a large number of cars, bad roads with attendant traffic jams locally known as "go-slow", vehicle traffic was considered the greatest pollutant.

Protection against the health hazards of noise is nonexistent in most African countries. It is hoped that this study will be extended to other African countries with similar problems. The World Health Organization is urged to include noise investigation in its program on environmental health for the African zone to make "Health for All by the year 2000" a meaningful reality.

ACKNOWLEDGEMENTS

The author acknowledges the assistance of the 1984 graduate majors in Audiology, University of Ibadan in obtaining part of the measurements, and my wife Bernadette for typing the manuscript.

REFERENCES

3. C.A. Bakare, "The Effects of Industrial Noise in Nigeria" (in press).
ALCRAFT NOISE ABATEMENT CORRESPONDING TO TOWN- AND REGIONAL PLANNING?

Lothar G.S. Prang

*ASI* 
Büro für Architektur, Städtebau, Infrastruktur
D-4044 Kaarst 2, Kubertusstraße 10
West Germany (FRG)

Introduction

Notable results have been experienced after several years of research and practice to reduce the aircraft noise impact. But the annoying situation in the vicinity of major airports and the sensibility of the public opinion against a/c-noise have not yet reached tolerable proportions with respect to the importance of the coordination of progressive town-planning and fast-developing civil aviation. This is substantially evident by the historic events: in the beginning of civil aviation after WW II the rather small airfields were constructed far outside the cities they had to serve, but taking into account the suspected dangerous procedures of take-off and landing, maintenance, and last not least aircraft noise.

After WW II civil air traffic has made a big jump towards mass transportation with all the pros and cons one has to deal with today. In the increased trade and commerce, the growing population of the big cities and consequently the enormous demands of land for industry and housing, the existing airfields reluctantly became integrated - but often disliked - parts of the urban area planning. However, in many cases the additional impact of airport movements beyond the airport boundaries was not taken into adequate consideration by the responsible planning authorities. Moreover, the built-up areas moved towards the airport by practical intentions, thus coincidentally increasing the imminent conflict situation of different interests. (Fig.1).

The technical progress of aeronautics, the introduction of jet-propulsion, and the necessary extension of the airports were not realized in time or even carelessly neglected in spite of early advice and warning of experts, e.g. the J.H. Doolittle Report in 1952. The usual but delayed administrative procedures or compromising policy to settle the perceivable controversial affairs proved to be exceptionally ineffective, and generally evoking persistent juridical issues, often with unsatisfactory consequences on either side.

Planning and Legal Provisions

Town- and Regional Planning are precisely founded on administrative and juridical procedures, striving for impartial solutions to the benefit of public affairs. Therefore town- and regional planning activity has to include necessary measures for the present but also for a comprehensive future, whilst legal claims are in principle bound to actual facts. That's why the coordination of a/c noise abatement with town- and regional-planning, adequate extension of airports, and the management of increasing air traffic is so difficult.

Many airports and communities are facing the same a/c-noise abatement problems, but notwithstanding the similarity of impending details there are hardly means of equivalent solutions to be met: the special local and regional realities are in most cases very different, and demand particular actions.

In the F.G. the legal Act Against Air Noise was enacted in 1971, much too late to prevent or to remedy existing and preceding facts. The Act defines a noise reduction area outside the airport boundaries by two protecting zones, I: $L_{eq}$ more than 75 d$\text{B(A)}$, II: $L_{eq}$ more than 67 d$\text{B(A)}$. The construction of hospitals, schools and other sensitive objects is prohibited in both zones. Residential premises are permitted in Zone II. Some exceptions small prevent unintended disadvantage or injustice, e.g. for houses built prior to the date of the Act. Subsidies can be granted for the improved insulation of homes, up to D: 150,- /qm flat area. The airport operator is liable for the compensation. (Fig.2).
During the past two decades several attempts have been made in the FRG to remove airports to another area. The Hamburg-Kaiserswerth project failed after enduring and expensive preliminaries. The Munich II-project has reached the initial stage of construction after lengthy private objections and claims of adjacent land owners.

The often repeated proposal to close an airport, and to construct a new one in a more suitable location, and to give these projects high priority may be implemented in some favorable circumstances, but proved to be impracticable in densely populated regions. Even limited extensions of an airport exceed often the bounds of public acceptance, e.g. the Frankfurt turbulences have attracted worldwide attention.

Conclusions and Recommendations

The future task for the effective restoration, clearance and reconstruction of a huge area in the environment of an airport requires the courage and influence of independent politicians and capable planners. There is no doubt, that in many cases it shall be difficult to obtain an perfect or even ideal solution to eliminate the mistakes of the past period of twenty or more years. It is, however, an absolute requirement and public responsibility to start the necessary steps for the coordination of differing interests without delay. The time for the planning and realization should be rated for at least 20 years.

The planning concept has to be based on the expected capacity of the airport, in coordination with the commercial and environment development of the region. Main topics for the land use compatibility are flight safety and noise impact, generally decreasing with the increasing distance from the airport. It must be a clearly understood and accepted principle, that an airport is an important component of public affairs and has to become an integrated part of the regional infrastructure. (Fig.3).

by rail and road is an essential factor to facilitate the time-saving access to the terminal and loading area, and to reduce the requisite floor-space for parking lots. The reduction of individual car traffic is also an abatement of traffic noise and exhaust molestation. (Fig.4).

The airport authorities are responsible to initiate effective measures to reduce the annoyance of the adjacent inhabitants by the

- reduction of the a/c-noise inflicted area,
- clustering of the flight pathes in coordination with air traffic control agencies,
- introduction of less noisy approach-, landing and take-off-procedures,
- reduction of ground traffic by connection of the airport to the adjacent cities by adequate means of public transportation, etc., depending on local and regional conditions and facts.

The aircraft industry has made enormous efforts during the past years to diminish the a/c-noises and to conform with the recommendations of ICAO Annex 16. Additional improved results may be expected, but not in the equivalent extent as experienced during the past decade.

References:

Bull, Berunek and Newman Inc.: Land Use Planning relating to aircraft noise. New York 1964/65

L. Frangi: Einordnung der Luftfahrtinfrastruktur in Städtebau und regionalplanung. Jacob 1979

Papers: several lectures of the author, e.g. 1956 Scheveningen, II. Europ. aeronautic congress, 1974 Kiel, CCLR Congress 1994 Honolulu, Inter-noise 84
1999 Munich, "

Fig. 3

The initial element for the planning authorities is the analysis of the present suburban fringes "moving towards the airport", how to stop this trend or to alter it by directing it to less sensitive areas. The improvement of public transport to the airport

Fig.4

ICAO 1986

ASLI 12

Toronto
1986
MEASUREMENT OF NOISE EMITTED BY DIFFERENT TRAINS IN DIFFERENT STATIONS OF MADRID UNDERGROUND

J.S. Santiago, J. Pons, E. Mateos and E. Costa

Instituto de Acústica, c/ Serrano 144, 28006 Madrid, Spain.

INTRODUCTION

With the intention of improving the acoustical conditions of the stations of Madrid's Underground, it was performed a series of measurements of the noise produced by the movements of the trains in several stations of different shape and size, that can be considered as representative of the majority of the existing ones. All the stations are of the double track type. For the measurements, three different types of trains were used: the oldest in circulation in the network, called "classic", and two other types of modern fabrication types 5000 and 2000, being this one the newest, still in essay in the lines, and with few units in circulation. The characteristics of the cars, both tractors and trailers, differ substantially.

The results obtained have been applied to a simple model of train circulation that has allowed to evaluate, for the different types of stations studied, the equivalent continuous sound level and the noise doses of exposure of the personnel working in the stations.

NOISE MEASUREMENTS

For the noise measurements, two one inch condenser microphones were installed, one at the center of each one of the platforms of the station, 1.2 m above the floor level; the noise signal was recorded in instrumentation magnetic tape recorders, and later analysed, both in frequency bands in a real time analyzer and statistically. Simultaneous recording in both microphones were made for the periods of time corresponding to a typical maneuver of the trains in the stations, that is, arrival of the trains, stop, train doors opening, acoustical warning signal, door closing and exit of the train from the station. In Table I are presented the main characteristics of the measured stations (size and shape) and the type of the train measured in each one.

TABLE I.- Characteristics of the stations, type and number of cars of the trains.

<table>
<thead>
<tr>
<th>Station</th>
<th>Width(m)</th>
<th>Length(m)</th>
<th>Type</th>
<th>Trains</th>
<th>Cars</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13.82</td>
<td>90.28</td>
<td>Straight</td>
<td>Classic</td>
<td>6</td>
</tr>
<tr>
<td>B</td>
<td>11.05</td>
<td>60.10</td>
<td>Straight</td>
<td>Classic</td>
<td>4</td>
</tr>
<tr>
<td>B</td>
<td>11.85</td>
<td>60.18</td>
<td>Straight</td>
<td>7000</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>13.82</td>
<td>60.30</td>
<td>Curved</td>
<td>Classic</td>
<td>4</td>
</tr>
<tr>
<td>C</td>
<td>13.82</td>
<td>60.30</td>
<td>Curved</td>
<td>2000</td>
<td>4</td>
</tr>
<tr>
<td>D</td>
<td>11.85</td>
<td>60</td>
<td>Curved</td>
<td>Classic</td>
<td>4</td>
</tr>
<tr>
<td>E</td>
<td>11.85</td>
<td>59.95</td>
<td>Straight</td>
<td>Classic</td>
<td>4</td>
</tr>
<tr>
<td>F</td>
<td>15.45</td>
<td>114.51</td>
<td>Straight</td>
<td>5000</td>
<td>6</td>
</tr>
<tr>
<td>G</td>
<td>15.45</td>
<td>115</td>
<td>Straight</td>
<td>5000</td>
<td>2</td>
</tr>
</tbody>
</table>

The frequency analysis have been made splitting the manoeuvres in two parts, arrival of the train and door opening, and door closing and exit, eliminating the acoustic warning signal. On the contrary, in the statistical analysis for the calculations of Leq and noise doses, the warning signal has been included.

In Figure 1 are presented three frequency analysis, in 1/3 octave bands, of the noise produced by the entrance of the three types of trains. The measurements were made during non service hours, with only five persons in the stations. In Figure 2 are presented the frequency spectra corresponding to the exit.

![Figure 1](image1.png)

Figure 1.- 1/3 octave band analysis of the noise produced by the entrance of the trains along the far track. — Classic train; --- 2000 train; ---- 5000 train.

![Figure 2](image2.png)

Figure 2.- 1/3 octave band analysis of the noise produced by the exit of the trains along the far track. — Classic train; --- 2000 train; ---- 5000 train.

EVALUATION OF THE PERCEIVED NOISE

To evaluate the noise perceived by the service personnel working in the stations (station masters, etc...), it was taken as a basis a simple model of circulation of trains that allows, knowing the number of trains circulating in a given moment in the line and the mean time of a round trip, to calculate the frequency of passage for each station, and hence the number of trains passing along each station, in each sense, within a time interval.

A statistical analysis of the noise recordings was made, covering the time intervals in which the noise of the trains was above the background noise in the station (including noise produced by the trains inside the tunnel). The equivalent continuous sound levels, Leq, and the percentiles L90, that were
adopted as the values of the background noise levels that would exist in the stations during the working hours, were obtained.

The underground system functions during twenty hours (6 a.m. to 2 a.m.), the personnel in the stations working in three shifts; so, a period of four hundred minutes has been taken as time of exposure. For those 400 minutes have been calculated the Leq in each one of the stations.

The working places in the stations are situated inside a room placed in one of the platforms, with large glass windows and also a glass door that is usually open; the sound reduction from the outside has been evaluated to be of 3 dBA in the mean. The Leq corresponding to the circulation of the trains along the near track and the far track have been computed separately.

The noise of the trains circulating along the near and far tracks, respective of the position of each microphone differ in an amount of the order of more than one dBA, being the engines of the "near train" shielded by the platform.

For the calculation of the perceived noise doses inside these cabins, a reference of a Leq of 85 dBA during 8 hours has been taken. In Table II are presented the calculated noise doses for the three shifts in the different stations.

**TABLE II.** Parameters used for the calculation of noise exposure and equivalent continuous sound levels obtained for each shift.

<table>
<thead>
<tr>
<th>Station</th>
<th>Train</th>
<th>N₁</th>
<th>N₂</th>
<th>N₃</th>
<th>d(δ)</th>
<th>Leq₁</th>
<th>Leq₂</th>
<th>L₉₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Classic</td>
<td>251</td>
<td>263</td>
<td>72</td>
<td>73.3</td>
<td>86.7</td>
<td>87.7</td>
<td>66.9</td>
</tr>
<tr>
<td>B</td>
<td>Classic</td>
<td>185</td>
<td>310</td>
<td>87</td>
<td>68.9</td>
<td>84.4</td>
<td>85.1</td>
<td>64.6</td>
</tr>
<tr>
<td>B</td>
<td>2000</td>
<td>185</td>
<td>310</td>
<td>87</td>
<td>53.3</td>
<td>82.5</td>
<td>84.4</td>
<td>64.6</td>
</tr>
<tr>
<td>C</td>
<td>Classic</td>
<td>185</td>
<td>310</td>
<td>87</td>
<td>67.1</td>
<td>84.1</td>
<td>84.7</td>
<td>66.4</td>
</tr>
<tr>
<td>C</td>
<td>2000</td>
<td>185</td>
<td>310</td>
<td>87</td>
<td>54.1</td>
<td>81.6</td>
<td>82.2</td>
<td>66.4</td>
</tr>
<tr>
<td>D</td>
<td>Classic</td>
<td>242</td>
<td>270</td>
<td>89</td>
<td>75.1</td>
<td>86.9</td>
<td>87.4</td>
<td>64</td>
</tr>
<tr>
<td>E</td>
<td>Classic</td>
<td>242</td>
<td>196</td>
<td>03</td>
<td>82.4</td>
<td>86.1</td>
<td>85.1</td>
<td>69.4</td>
</tr>
<tr>
<td>F</td>
<td>5000</td>
<td>181</td>
<td>173</td>
<td>76</td>
<td>83.1</td>
<td>81.6</td>
<td>82.7</td>
<td>66</td>
</tr>
<tr>
<td>G</td>
<td>5000</td>
<td>238</td>
<td>100</td>
<td>70</td>
<td>64.7</td>
<td>79.2</td>
<td>81.1</td>
<td>60.4</td>
</tr>
</tbody>
</table>

N₁ - Number of trains for shift 1 (400 minutes) passing along the station.

d - Mean duration of train manoeuvre in the station.

Leq₁ - Leq in dBA corresponding to a train manoeuvre in the station along near track relative to the personnel cabin.

Leq₂ - Id. along far track.

L₉₀ - Background level in the station in dBA.

In Table III are presented the total Leq for each shift, corresponding to the train passage along both tracks and the quiet periods, in the platform; also the noise doses received inside the personnel cabins.

**TABLE III.** Leq in the station platforms and total noise doses (in %) inside the cabins, per shift.

<table>
<thead>
<tr>
<th>Station</th>
<th>Train</th>
<th>Leq₁</th>
<th>Leq₂</th>
<th>Leq₃</th>
<th>D₁</th>
<th>D₂</th>
<th>D₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Classic</td>
<td>86.1</td>
<td>86.2</td>
<td>80.8</td>
<td>53.7</td>
<td>55</td>
<td>15.8</td>
</tr>
<tr>
<td>B</td>
<td>Classic</td>
<td>82.1</td>
<td>84.3</td>
<td>78.9</td>
<td>21.4</td>
<td>35.5</td>
<td>10.2</td>
</tr>
<tr>
<td>B</td>
<td>2000</td>
<td>79.7</td>
<td>81.9</td>
<td>76.5</td>
<td>12.3</td>
<td>20.4</td>
<td>5.9</td>
</tr>
<tr>
<td>C</td>
<td>Classic</td>
<td>81.6</td>
<td>83.8</td>
<td>78.4</td>
<td>19.1</td>
<td>31.6</td>
<td>9.1</td>
</tr>
<tr>
<td>C</td>
<td>2000</td>
<td>78.3</td>
<td>80.4</td>
<td>79.6</td>
<td>8.9</td>
<td>14.5</td>
<td>12</td>
</tr>
<tr>
<td>D</td>
<td>Classic</td>
<td>87.2</td>
<td>86.4</td>
<td>81.7</td>
<td>69.2</td>
<td>57.5</td>
<td>19.5</td>
</tr>
<tr>
<td>E</td>
<td>Classic</td>
<td>84.9</td>
<td>84</td>
<td>80.4</td>
<td>40.7</td>
<td>33.1</td>
<td>14.5</td>
</tr>
<tr>
<td>F</td>
<td>5000</td>
<td>80</td>
<td>79.8</td>
<td>76.4</td>
<td>13.2</td>
<td>12.6</td>
<td>5.8</td>
</tr>
<tr>
<td>G</td>
<td>5000</td>
<td>78.4</td>
<td>74.7</td>
<td>73.2</td>
<td>9.1</td>
<td>3.9</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Leq₁ - Total Leq in dBA for shift 1 (400 minutes)

D₁ - Noise dose for shift 1, in %

**COMMENTS**

Stations B and C belong to the same line and the measurements were made with the same trains and drivers; the comparison between the noise in straight and a curved station shows the influence of the speed of the train when entering and leaving the stations, both for classical and 2000 trains.

The small differences among the Leqs produced by classical and 2000 trains, although the running of the latter is less noisy, is due to the higher level of the acoustic warning signal emitted by these trains.

The 5000 trains, apart from being relatively silent, circulate in modern lines with better acoustic treated stations (F and G), with volumes twice the volumes of the older stations (A to E). The noise doses received by the personnel inside the cabins are well below the recommended values; only in four cases, in first two shifts in stations A and D, the 100% dose would be reached in the station platforms.

**ACKNOWLEDGMENTS**

The authors express their acknowledgment to the Compañía del Metro, Madrid for the permission granted for the measurements; special gratefulness is due to Mr. Navarro of the said company for his help and informations.
RURAL DETECTABILITY LIMITS FOR ARMY MATERIEL

G. R. Garthner and J. T. Kalb
U.S. Army Laboratory Command
Human Engineering Laboratory
Aberdeen Proving Ground, MD, 21005-5001, USA

INTRODUCTION

The U.S. Army requires design limits to prevent rural detectability of materiel at various ground-to-ground distances. These limits are to be specified in 1/3-octave band levels which should not be exceeded at given measurement distances to prevent detectability at corresponding listener distances.

Recently, new data (1977) have prompted an examination of those factors affecting the propagation and detection of sound with the goal of providing a viable detectability standard. There is no new information on sound attenuation caused by ground effect; on human ability to distinguish sounds from background noise; on variation of background noise with location; and on prediction of atmospheric absorption. This paper will describe the procedure used to develop such a limit which will serve as a goal for the design of materiel like generators, vehicles, or pumps which have a detectability requirement.

PROCEDURE

We determined that it would be appropriate to provide the user of this standard with a choice of two limits. The first would be based upon the quietest noise level which would be expected at a listener location far from heavily traveled highways and communities. The second would be based on noise levels which are typically found in rural areas closer to mammae noise.

The factors considered in determining the limits are: background noise, threshold of hearing, psychoacoustic factors, geometric spreading, atmospheric absorption, ground effect, atmospheric turbulence, refraction due to wind and temperature gradients, barriers, and foliage.

Since it is not practical to provide detectability limits for all of the permutations of each factor above, only certain factors were selected for the formulation of this limit. In addition, for those factors included, certain conditions were selected as being typical or most likely to be encountered under many situations (like type of ground surface, or temperature). The factors and conditions selected were chosen to err on the side of conservatism. For example, during windy daytime conditions, material will probably be much less detectable than predicted by the limit; however, during still nighttime conditions, material may be slightly more detectable than predicted by the limits.

The following are the factors which were included and the conditions which were assumed.

PARAMETERS AFFECTING DETECTION OF SOUND

Background noise is probably the single most important factor for determining the detectability of a sound. We have therefore based the selection of the two limits (critical and typical) provided in this standard, upon two residual daytime background noise levels found in the U.S., as reported by the Environmental Protection Agency (1977). These are the noise levels found at the North Rim of the Grand Canyon, which was the quietest area they found, and the noise levels found in typical rural farm areas.

Figure 1 shows the range of 1/3-octave band levels found at these locations with the lower limit of each range being the residual noise levels. The difference between these two levels is representative of the different contributions to noise made by man and machine at two different distances. Computations have shown that these the noise levels are typically 4 and 16 kilometers from communities and heavily used highways.

![Background noise assumed for the standard.](image)

Psychoacoustic factors play an important role in the ability of people to detect sounds. The threshold of hearing used for this standard is that published in ISO 8226. This represents the hearing sensitivity of young, normal, non-noise-exposed individuals. The 200 ms temporal integration time of the auditory system is taken into account by requiring that measurements be made with fast meter damping and that the maximum meter deflection be the recorded value. A listener's bias or certainty in making a response has not been accounted for by including the theory of signal detection (TSD). The following TSD parameters have been assumed in this standard: the listener's hit probability is 50 percent, false alarm rate is 1 percent, and the listener is 40 percent as efficient as an ideal observer. These parameters are characteristic of highly motivated listeners.

PARAMETERS AFFECTING SOUND PROPAGATION

Propagation of sound through the atmosphere is controlled by a number of phenomena each dependent upon different factors. Those factors included in this standard and the assumed conditions for each are as follows.

Geometric spreading is uniform at all frequencies and causes sound in the free field to decrease at 6 dB per doubling of distance. For an accurate prediction of geometric spreading, the noise measurement must be made in the free field, which begins at a nominal distance of 3-5 times the major dimension of the source.

Atmospheric absorption is dependent upon distance, frequency, relative humidity, temperature and, to a very small extent, atmospheric pressure. This excess attenuation, which is in addition to that due to geometric spreading, is caused by the sound waves losing energy to the vibrating oxygen and nitrogen molecules. The assumed nominal conditions for this standard are 15 degrees C and 70 percent relative humidity, with the computed atmospheric absorption being based upon ANSI S1.26.

Ground effect produces significant attenuation at frequencies ranging from 250 to 1000 Hz. This
Effect is due to the interaction of the reflected sound wave off the ground and the direct sound wave, and is affected by the turbulence of the atmosphere near the ground. The parameters involved are frequency, source and listener height, distance, ground surface, and nonhomogeneity of the atmosphere. Surfaces such as grass can provide excess attenuation up to 20 dB when both the source and the listener are near the ground; other surfaces produce different degrees of excess attenuation (Figure 2). This effect is significantly reduced if either is greater than 10–15 meters above the ground. The assumed conditions for this standard are that the ground surface is grass, source and listener heights are 1.2 meters, and calm atmospheric conditions exist. Noise measurements will be made at 1.2 meters above a uniform, flat, grass surface.

![Diagram of ground effect over three surfaces.](image)

**Table 1. Limits, in dB, for critical detectability**

<table>
<thead>
<tr>
<th>Freq (Hz)</th>
<th>10</th>
<th>30</th>
<th>100</th>
<th>300</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>55</td>
<td>62</td>
<td>61</td>
<td>71</td>
<td>68</td>
<td>74</td>
<td>81</td>
<td>88</td>
</tr>
<tr>
<td>60</td>
<td>46</td>
<td>56</td>
<td>55</td>
<td>65</td>
<td>62</td>
<td>69</td>
<td>76</td>
<td>85</td>
</tr>
<tr>
<td>80</td>
<td>39</td>
<td>49</td>
<td>50</td>
<td>60</td>
<td>57</td>
<td>64</td>
<td>73</td>
<td>84</td>
</tr>
<tr>
<td>100</td>
<td>33</td>
<td>42</td>
<td>45</td>
<td>53</td>
<td>51</td>
<td>61</td>
<td>72</td>
<td>86</td>
</tr>
<tr>
<td>125</td>
<td>25</td>
<td>34</td>
<td>41</td>
<td>52</td>
<td>50</td>
<td>61</td>
<td>74</td>
<td>87</td>
</tr>
<tr>
<td>150</td>
<td>21</td>
<td>30</td>
<td>37</td>
<td>50</td>
<td>50</td>
<td>63</td>
<td>77</td>
<td>93</td>
</tr>
<tr>
<td>200</td>
<td>15</td>
<td>25</td>
<td>35</td>
<td>50</td>
<td>52</td>
<td>67</td>
<td>74</td>
<td>92</td>
</tr>
<tr>
<td>250</td>
<td>12</td>
<td>20</td>
<td>27</td>
<td>34</td>
<td>35</td>
<td>44</td>
<td>60</td>
<td>88</td>
</tr>
<tr>
<td>315</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>26</td>
<td>30</td>
<td>38</td>
<td>59</td>
<td>98</td>
</tr>
<tr>
<td>400</td>
<td>7</td>
<td>10</td>
<td>13</td>
<td>18</td>
<td>20</td>
<td>26</td>
<td>38</td>
<td>59</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>18</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>630</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td>16</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td>700</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>15</td>
<td>22</td>
<td>32</td>
</tr>
<tr>
<td>800</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>15</td>
<td>23</td>
</tr>
</tbody>
</table>

**Table 2. Limits, in dB, for typical detectability**

<table>
<thead>
<tr>
<th>Freq (Hz)</th>
<th>10</th>
<th>30</th>
<th>100</th>
<th>300</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>53</td>
<td>62</td>
<td>61</td>
<td>71</td>
<td>68</td>
<td>74</td>
<td>81</td>
<td>88</td>
</tr>
<tr>
<td>60</td>
<td>46</td>
<td>56</td>
<td>55</td>
<td>65</td>
<td>62</td>
<td>69</td>
<td>76</td>
<td>85</td>
</tr>
<tr>
<td>80</td>
<td>39</td>
<td>49</td>
<td>50</td>
<td>60</td>
<td>57</td>
<td>64</td>
<td>73</td>
<td>84</td>
</tr>
<tr>
<td>100</td>
<td>33</td>
<td>42</td>
<td>45</td>
<td>53</td>
<td>51</td>
<td>61</td>
<td>72</td>
<td>86</td>
</tr>
<tr>
<td>125</td>
<td>25</td>
<td>34</td>
<td>41</td>
<td>52</td>
<td>50</td>
<td>61</td>
<td>74</td>
<td>87</td>
</tr>
<tr>
<td>150</td>
<td>21</td>
<td>30</td>
<td>37</td>
<td>50</td>
<td>50</td>
<td>63</td>
<td>77</td>
<td>93</td>
</tr>
<tr>
<td>200</td>
<td>15</td>
<td>25</td>
<td>35</td>
<td>50</td>
<td>52</td>
<td>67</td>
<td>74</td>
<td>92</td>
</tr>
<tr>
<td>250</td>
<td>12</td>
<td>20</td>
<td>27</td>
<td>34</td>
<td>35</td>
<td>44</td>
<td>60</td>
<td>88</td>
</tr>
<tr>
<td>315</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>26</td>
<td>30</td>
<td>38</td>
<td>59</td>
<td>98</td>
</tr>
<tr>
<td>400</td>
<td>7</td>
<td>10</td>
<td>13</td>
<td>18</td>
<td>20</td>
<td>26</td>
<td>38</td>
<td>59</td>
</tr>
<tr>
<td>500</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>18</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>630</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>12</td>
<td>16</td>
<td>24</td>
<td>36</td>
</tr>
<tr>
<td>700</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>15</td>
<td>22</td>
<td>32</td>
</tr>
<tr>
<td>800</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>9</td>
<td>13</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>1000</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>10</td>
<td>15</td>
<td>23</td>
</tr>
</tbody>
</table>

**Figure 2. Ground effect over three surfaces.**

Refraction due to wind and temperature gradients can cause sound waves to either bend toward or away from the ground, affecting detection at distances greater than 50 meters. During the day, when temperatures are warmer near the ground, sound waves bend upward causing a shadow zone with up to 25 dB loss. At night, temperatures are typically cooler near the ground, and sound waves are bent downward with a resultant sound increase of up to 3 dB above neutral conditions. Likewise, compared to still conditions, refraction due to wind can cause attenuation of up to 25 dB for sound traveling into the wind and a decrease in attenuation of up to 3 dB when traveling with the wind. For this standard, neutral wind and temperature conditions are assumed. Barriers and foliage are assumed to be absent and sparse, respectively, for this standard.

**Computation of the Limits**

The limits for conformance to this standard were computed for the specified measurement distances by using the preceding factors. The noise limits for material are those levels, for each 1/3-octave band, which produced a 0 dB signal-to-noise ratio for the two assumed background noise spectra at the detectability distance. The following tables provide the limits for critical and typical detectability. They show the 1/3-octave band levels that are not to be exceeded at the measurement distance specified, for detectability not to take place at various distances.

**References**


SOUND EMISSION FROM RIFLE SHOOTING RANGES

U. J. Kurze, L. Schreiber
Miller-BSM, Robert-Koch-Straße 11, D-8033
Planegg/München, West Germany

INTRODUCTION

Noise from shooting ranges is different from most industrial and traffic noises because of its impulsiveness, large peak pressure values, and the possible combination of point and line sources, both having pronounced directivities. Under contract of the Federal Environmental Agency (Umweltbundesamt, Berlin), we have carried out measurements on the emission and propagation of sound from various shooting ranges in order to provide a data base and a thorough background for the development of a catalogue of noise control measures. This paper reports our findings on the question whether the emission of shooting noise has any peculiarities that may not allow for applying the standard procedures of noise control, comprising the description of sound sources in terms of sound power level, frequency content, and directivity, and the definition of noise control measures close to the sources, such as absorptive linings of reflecting blinds, earth berms and sound barriers.

SOUND MEASUREMENTS

The cross section of a Type B military shooting range is shown in Fig. 1 together with the microphone positions located aside from the rifleman and at distances of 50 and 100 m from the muzzle. Other bullet paths, rifleman standing

lying

25 m

8 m

25 m

32 m

Fig. 1 Schematic cross section of 200 m shooting range with microphone positions

positions were 5 and 10 m behind the rifleman. While the Type B range has no blind except from the target area, measurements have also been carried out on Type A and C ranges with various blinds along the total length of 300 m. 1/4" condenser microphones, short cables and preamplifiers with 120 V biasing voltage have been used together with a multichannel FM tape recorder and a dual channel signal memory recorder. The upper frequency limit of the equipment was about 50 kHz.

During the repeated firing of single shots from the standard G3 rifle, two types of signals have been observed, namely the muzzle report aside from and behind the rifleman and the ballistic shock wave emanating from the bullet. Examples given in Fig. 2 show an inversion of the preamplifier, so that a negative voltage corresponds to a positive pressure and vice versa. The typical duration of the muzzle report is 1.2 ms, corresponding to a spectral max-

2.27 ms

3.4 ms

Ballistic shock wave and ground reflection

2.5 m from bullet path, 50 m from lying rifleman

Muzzle report and ground reflection,

8 m aside from standing rifleman

Fig. 2 Signal records

mum at 800 Hz, and the N-shaped ballistic wave has a typical duration of 0.25 ms corresponding to 4 kHz. Both signals are followed by reflections from the grass covered ground. The difference in time between direct sound and ground reflections (or reflections from earth berms, single blinds etc.) and between the ballistic shock wave and the muzzle report (at some points in front of the rifleman) could be satisfactorily related to the geometry and to an average bullet velocity of 725 m/s. Multiple reflections between blinds show a reverberant field without the possibility to identify individual contributions.

EVALUATION

Recordings of 10 or more repeated free field sound measurements have been analysed in terms of the average peak sound pressure, the average maximum pressure gradient and the average signal duration. From linear theory, one would expect signal durations (and shapes) that are independent of distance, and amplitudes that decay with 1/r from a point source, corresponding to a level reduction of 6 dB per doubling of distance, and with 1/r² from a line source, corresponding to 3 dB/dl. Any increase in signal duration and excess attenuation at high sound levels indicates nonlinearities. As well known from theory 1/r, the duration of a cylindrical shock wave increases at some distance from the source as r²/λ. In fact, we observed such an increase with r² in the range from 2.5 to 25 m (Fig. 3). At a distance of 32 m, however, where the microphone had been located behind the earth berm, diffraction has a major effect on the signal shape. The decay of the shock front of a cylindrical or conical wave with distance as r⁻³/². Again, the evaluation of our measured data shown in Fig. 4 is quite consistent with the non-linear theory within the range of free field sound propagation. The shielding of the earth berm then provides for an additional attenuation between 25 and 32 m. It must be emphasized that rifle shooting ranges always contain protective earth berms. Therefore, the reduction of the stronger decaying shock...
wave to a more sinusoidal wave with linear propagation behaviour in the range of amplitudes below 120 dB is a typical feature that must be considered for environmental prediction purposes. For a path length difference characteristic for the shielding of the earth berm $z = h^2/2a = 9/50$ m and a typical frequency of 2.5 kHz at 30 m, the excess attenuation is $\Delta L = 10 \log (3 + 20 s/0.07) = 17$ dB. Thus, the level of 153 dB given in Fig. 4 may be reduced to 138 dB and the dependence on distance from -15 to -10 log $r$ in order to apply a linear prediction model beyond the earth berm.

Theoretically, the shock amplitude of a spherical wave decays inversely proportional to $r \sqrt{\ln r/h}$, where $h$ is the source radius of the shock wave. The measured data plotted in Fig. 5 consistently show a decrease of the peak sound pressure level of the muzzle report, which slightly exceeds -20 log $r$. Data measured by Holtrup /2/ within 2 m from the muzzle are somewhat higher than our data measured in the range from 2.5 to 100 m. Rather pronounced is the directivity. In the shooting direction, the peak level is roughly 6 dB higher, and in the perpendicular direction, it is consistently lower than the averaging function 161 dB - 20 log $r$. The peak levels of ground reflections were found 3 dB lower for a standing rifleman and at least 6 dB lower for a lying one. This may be attributed to the different angles of reflection. Evaluation of the peak pressure gradient $dP/dt$ in the range from 2.5 to 25 m precisely showed a $1/r$ power law.

$$L_{\text{AFmax}} = 138 + 10 \log \left( \frac{S_{\text{A}}}{S_{\text{B}}} \right) + 0.23 \log \left( \frac{1}{S_{\text{A}}} \right)$$

The draft of a VDI guideline for the assessment of shooting noise is based on maximum values of $A$-weighted sound levels measured with the time constant FAST. Such data have been evaluated and are shown in Fig. 6. For microphone positions in front of and aside from the rifleman, they are well approximated by a formula employing the power laws $1/r$ and $1/r^2$ for the linear propagation of spherical and cylindrical waves, respectively. At a reference distance of 1 m, the level of the muzzle report is 5 dB higher than the level of the ballistic shock wave. The level of 138 dB(A) is consistent with the Swiss prediction scheme /3/ for the muzzle report (145 dB(A), accounting for an $dA$ directivity between 0° and 90°), and so is the level of the ballistic shock wave (121 dB(A) at 1 m). In fact, the level of the ballistic shock wave is also well supported by the following formula derived from conversions of the peak level, the duration of the N-wave, a ground reflection 3 dB below the direct sound, and a minor correction for the $A$-weighting:

$$L_{\text{AFmax}} = L_{\text{peak}} + 10 \log \left( \frac{T}{125} \right) + 10 \log \left( \frac{1 + 0.5}{\Delta A} \right)$$

For $r = 2.5$ m, $L_{\text{peak}} = 148.5$ dB, $T = 0.19$ ms and $\Delta A = 0.5$ dB, one finds $L_{\text{AFmax}} = 118$ dB(A) or 122 dB(A) - 10 log $r$.

CONCLUSION

In spite of high source levels, a linear model has been found sufficient to describe the sound emission from rifle shooting ranges, accounting for both the muzzle report and the ballistic shock wave and including the shielding of sound by protective earth berms. The data are consistent with the Swiss prediction scheme presented after completion of our field investigations.

/2/ G. Holtrup, Results from the measurement program at Mainge, communication to the VDI Working Group B1-3745 (5.9.1985)
PEAK LEVEL FLUCTUATION OF IMPULSIVE NOISE OUTDOORS

L. Schreiber, U. J. Kurze
Müller-RM GmbH, Robert-Koch-Straße 11,
D-8033 Planegg/München, West Germany

INTRODUCTION

The draft WU guideline 3745 on the assessment of shooting noise suggests downwind measurements of the A-weighted SPL with time constant FAST. At least 10 independent data points must be measured for small level fluctuations less than 8 dB. The minimum number of measurements goes up to 30 for maximum permissible fluctuations of the observed levels between 13 and 15 dB. Such numbers ensure at 80 % probability that the energy mean level can be accurately determined within a range of ± 2 dB for a Gaussian distribution of energies. While a small number of repeated measurements is rather common and sufficiently accurate for extended industrial or traffic noise sources, it is the impulsiveness of shooting noise that results in much larger fluctuations and requires a comparatively large measurement effort. This paper describes some results of investigations carried out under contract from the Federal Environmental Agency (Umweltbundesamt), concerning the magnitude of level fluctuations and practical consequences for field measurements.

MEASUREMENTS

The plan view of a military rifle shooting range is shown in Fig. 1 together with measurement positions at the property line and in the neighbourhood. More than 30 shooting positions on Typ A, B, and C shooting ranges (different blinds) have been used for firing more than 2000 single shots during different times and days from G3 rifles with standard combat ammunition. A 4 m high wall on top of a 4 m high earth berm provides some protection of the neighborhood with the first house located at coordinate 700/200. The microphone Nos. 11 and 12 look over the wall, Nos. 1, 3, 5, 7, 9 and 10 are 4 m above ground, the others 5 to 8 m. From 1/4" and 1/2" microphones recordings have been made on digital signal recorders. The data have been evaluated in the laboratory in terms of $L_{A\text{max}}$.

OBSERVATIONS

The distribution of energy mean levels in the neighbourhood, averaged over 10 shots fired during 30 sec, is shown in Fig. 2. The time constant IMPULSE adds about 5 dB to the level $L_{A\text{max}}$ The data from different shooting positions (shooting ranges A, B, C, riffleman positioned right (R) or left (L), at 50 to 300 m from the target, lying (L) or standing up (S)) are spread over a range of 20 dB. Consequently, there is no way to restrict the level fluctuation to 15 dB unless controlled positions are used. Data observed for shooting from a particular position (CL 200 L) during 2 different days are plotted in Fig. 3. Again, these are energy mean levels. In spite of nominally equal "downwind" conditions, there are quite different results. As shown in Fig. 4, the fluctuation of individual measurements is rather small. In most cases, it was less than 8 dB. Simultaneous measurements don't exhibit a significant increase of level fluctuation with distance, as observed by Klein /1/ from tones warped in octave bands and from earlier

Fig. 2: Distribution of noise impact at MP 13, average data for different source positions

Fig. 3: Energy mean values of $L_{A\text{max}}$

Fig. 4: Cumulative distribution for different receiver positions
shock tube measurements /2/. A statistical analysis of the standard deviations, each evaluated from 10 successive shots, has been carried out for the nearest measuring point 11 and the furthest point 13. The cumulative distributions shown in Fig. 5 indicate wide ranges of data. However, at the near point 11, the fluctuations are significantly lower than those at point 13. There is a factor of 2 between the 50% values. At larger distances, the typical standard deviation of less than 2 dB is small compared to values of about 3 dB found from impulse sound measurements /2/ carried out over 20 min instead of 30 sec.

The range of level fluctuations observed by Holtrup /3/ on a circle of 400 m around a Type A shooting range during different seasons is plotted in Fig. 6. The data show some increase with increasing wind velocity but are rather similar for downwind and upwind conditions. Cross wind with varying direction might cause a fluctuating boundary of the shadow zone and, consequently, higher sound level fluctuations. Each data point is obtained from 3 series of measurements, carried out with 10 shots during 2 min and over a total time of 10 min. The average range of 12 dB level fluctuation may be roughly related to a standard deviation of 3 dB and, thus, is consistent with other long time measurements.

CONCLUSION

Measurements carried out over a short period of time under stable meteorological conditions result in small level fluctuations within a range of less than 3 dB requiring a reasonable number of 10 shots only. However, it is rather impossible to decide from such stabilities about favourable sound propagation conditions. The effect of sunshine or other causes for an excess attenuation may demand additional sound measurements in order to determine the true mean value of the "downwind" noise impact.

PROPOSAL

A shock tube has been developed that radiates a 1 ms sound impulse due to bursting of a pressurized membrane. This device can be set up outside of the shooting range where no shielding from earth bums and sound barriers is effective. Since the sound power of the source is known, one can calculate the free-field "downwind" sound pressure level. Measurements, preferably carried out with the time constant IMPULSE in order to exclude the unwanted contributions from nearby reflecting surfaces, are then compared to the calculated data, as shown by examples in Fig. 7. If the measured data are close to or exceed the calculated curve, the weather conditions are favourable for sound propagation. Fig. 7 indicates that the weather was useful for shooting noise measurements around noon, but not later in the afternoon.

/1/ Klein, M., Investigations on the sound propagation out of doors, LIS Report No. 42, 1983
/2/ Kürze, U.J., Shielding over large distances, FAP/RADAR '82, 359-362
/3/ Holtrup, G., Results from the measurement program at Mainz, Communication to the VDI Working Group BI-3745 (5.9.1985)
NOISE CONTROL AT CANADIAN NUCLEAR POWER STATIONS DURING THE DESIGN PHASE

K. M. Osman

Ontario Hydro, 700 University Avenue, Toronto, Ontario, M5G 1X6 Canada

INTRODUCTION

About one third of the electricity produced today in the Province of Ontario, Canada is generated from nuclear power stations. In addition, the share of nuclear power of the total generated capacity will increase in the years ahead. Currently there are six nuclear generating stations in operation in Ontario. All main stations have each four units ranging in output capacity from 500 to 750 MW per unit. The total estimated generation capacity from all nuclear units now stands at approximately 10,000 MW.

Except for the reactor which is an AEC (Atomic Energy of Canada Limited) design, most of other design work including the choice and placement of equipment is performed by the Corporation. Such work includes developing specifications and design requirements, performing tender evaluations, communicating with equipment manufacturers/suppliers and testing at the production shop and the installation site. Noise control work is integrated in the design process. For a major source of noise, a continual effort is deployed to abate the emanated noise.

Recent interest in the quality of the environment, both at the workplace and at the community, provided impetus to noise control measures during the design stage. Current provincial government regulations[1, 2] and Corporate directives[3, 4] reflect the importance given to this issue in Ontario. An orchestrated endeavor is being performed by the various areas of the corporation to achieve effective protection to the workers and to the community.

NOISE CONTROL AND DESIGN

The most effective measures to abate noise and to provide protection against hearing damage (workplace) or annoyance (community), are implemented at the noise source during the design phase. This statement conveys the current prevailing view regarding industrial noise control. The emphasis on the role of noise control through design activities is recognized in the applicable government regulations and Corporate directives. In such documents, engineering controls are usually given preference or priority over other types of controls (administrative or hearing protection).

There are certain circumstances or conditions that govern noise control activities in nuclear generating stations in Ontario, in particular:

(a) In each generating station, equipment type is identical for all units; and in addition, recently built generating stations have a similar layout (placement of equipment at the various elevations).

(b) Most of the time, the stations operate at a constant load condition, i.e. most equipment runs at the same load and speed.

(c) Sites for stations are always at a lake shore and the nearest community is typically 2 to 5 km away.

(d) In the powerhouse, the turbine/generator set is responsible for most of the 'background' noise across many parts of the station. Outdoors, the powerhouse is a major noise source at the community.

WORKPLACE NOISE CONTROL

The overall objective of workplace noise control is to ensure that workers do not incur any hearing damage due to high level noise exposure during their working hours. Another important objective is to minimize annoyance related to noise, hence provide basis for better work performance.

The applicable constraints prescribe exposure limits. They are as follows:

(a) Proposed Regulation

Limit worker's exposure to 90 dBA, Leq (equivalent sound level) based on 40 hour week; engineering controls are preferred.

(b) Ontario Hydro Corporate Program

Noise-exposed worker is identified if regularly exposed to a continuous time weighted average noise level, Leq, equal or above 85 dBA or impact noise between 135 to 140 dBA more than nine times a day or 140 dBA any time; engineering controls are given priority.

As a result of the above constraints and their implications, the approach taken now to control workplace noise in nuclear generating stations, is to address equipment responsible for workers' over-exposure as early as possible in the design phase. This can be achieved through the implementation of specific activities, namely:

(a) Prediction of workers' exposure for the different categories of workers' classifications (e.g. Operators, Mechanical Maintainers).

(b) Identification of specific equipment responsible for workers' over-exposure.

(c) Investigation of most effective engineering controls to abate high noise emissions from such equipment.

(d) Implementation of engineering controls and monitoring of their performance.

These activities are facilitated by certain conditions/circumstances that exist in nuclear generating stations (see above section on Noise Control and Design). Finally, the approach is selective and cost effective in that not all high noise emission equipment is to be addressed, only the ones that are responsible for workers' over-exposure.

COMMUNITY NOISE CONTROL

As part of the environmental requirements for nuclear generating stations, noise and vibration at the community are addressed starting with construction activities and continuing throughout the stations' life cycle (approximately 30 years). With respect to noise, the limits that govern the impact at the community, are given in a protocol signed with the Ontario Ministry of the Environment in 1981[5]. These limits pertaining to power generation can be summarized as follows:

(a) At the point of reception (30 m from a dwelling or a campsite), the maximum accumulative noise levels are: 55 dBA Leq 1 hour (power generation at heavy water plants).

(b) At the point of reception, the minimum measured Leq 1 hour of the ambient from 23:00 to 7:00 hours minus 5 dB subject to a lower limit of 35 dBA (transformers and
It should be noted that the above limits apply at the receivers' point which results in the inclusion of any meteorological or ground propagation effects. Therefore, the action taken here must include these effects. The approach can be summarized as follows:

(a) Identification of nearest communities outside the property lines of the site.
(b) Prediction of noise levels emitted from each major noise source (e.g., powerhouse, transformers).
(c) Prediction of noise levels at the identified communities assuming maximum noise impact (site activity, meteorological and propagation effects).
(d) Identification of sources requiring noise control measures.
(e) Implementation of engineering controls and monitoring of their performance.
(f) Carrying out of sound level surveys at the community before any site activities begin and during maximum impact times.

These activities usually span over a number of years and begin with the pre-construction surveys. Other surveys include maximum impact periods (e.g., outdoor activities in summer time). Sometimes, landscaping within the station boundaries produces measurable reduction in the noise level at the community (more than 3 dB) especially if it results in the interruption of the line of sight between the main noise source and the receiver.

EXAMPLE - DARLINGTON NGS

This is the newest nuclear generating station (NGS) in Ontario. The location is about 50 km east of Metropolitan Toronto on Lake Ontario. The station is under construction and has four units each rated at about 750 MW.

The noise control program for the station during the design stage has three phases: identification of noise sources (Phase I); selection of appropriate abatement measures (Phase II) and implementation and monitoring of performance of measures (Phase III). Phase I has been completed and currently Phase II is being pursued.

At the workplace a certain number of equipment/areas have been identified as noise sources and engineering controls are being investigated for some of them (e.g., Boiler Feed pumps and the area around the Primary heat transport pump motors). The table given below provides this information along with the predicted noise level in dBA.

Regarding noise levels at the community, the prediction for before and after operation is given in the site layout given below. The 'before' operation levels are based on several surveys conducted at various times and the 'after' operation levels are interpolations of measurements obtained from similar installation somewhere else in Ontario.

REFERENCES

(2) Model Municipal By-law under the Ontario Environmental Protection Act, August, 1978.
(3) Ontario Hydro Corporate Noise Control and Hearing Protection Program, March 19, 1984.
ETUDE DE PROPAGATION DU Bruit A GRANDE DISTANCE

M. Asselinneau, J.-G. Migmonen

Laboratoire d’acoustique et Centre de recherches en aménagement et en développement, 1624 pavillon P.-A.-Savard, Université Laval, Ste-Foy, Québec, GIK 7P4

Le bruit des transformateurs électriques constitue une source acoustique dont l’émission est permanente et le niveau sensiblement constant en fonction du temps. Du fait de son contenu en basses fréquences, ce bruit est également peu sensible à l’absorption moléculaire. Ces caractéristiques en font donc un sujet de choix pour une étude de propagation sur des distances supérieures au kilomètre.

L’étude présentée ici a été réalisée sur une période d’un an, autour d’un poste de transport 735 KV (poste Jacques-Cartier, région de Québec) situé en milieu naturel (semi-boisé). Les voies d’accès existantes ont permis des mesures selon trois axes géographiques d’une longueur de 2 km. Sur ce site, se trouvait également un ensemble de capteurs météorologiques fournissant, selon un cycle horaire, la température (mesurées à différentes hauteurs au-dessus du sol: 3 m, 73 m et 137 m), la vitesse et la direction du vent (6 m et 73 m) et l’humidité relative près du sol.

Tous les appareils de transformation du poste – du type ONAN (huile et air non forcées dans les radiateurs) – ont fait un préalable l’objet d’une mesure de leur puissance acoustique selon la norme NEMA, ainsi que d’une analyse spectrale au 1/3 d’octave; cette dernière ayant mis en évidence la nette prépondérance de la bande 125 Hz. La puissance acoustique temporairement totale devait être théorique permettre des mesures significatives d’atténuation sur des distances de l’ordre de 2.5 km.

Les distances de mesure au long des trois axes de propagation ont été fixées à 0 (clôture), 230, 500, 750, 1000, 1500, 1750 et 2000 m. Pour l’ensemble des 24 parties de mesure, les relevés réalisés ont été les suivants (2286 relevés):
- mesure du niveau moyen en dB(A) en l’absence de toute perturbation du bruit analysé autre que naturelle;
- mesure du niveau moyen dans la bande au 1/3 d’octave de 125 Hz; étant donné les fluctuations instantanées souvent présentes, ce résultat a été noté à partir des minimum et maximum obtenus;
- enregistrement graphique à 125 Hz d’une ondulation ou fluctuation intéressante (généralement entre 750 ou 1750 m, voir fig. 1).

48 dossiers de jour et 47 dossiers de nuit ont été ainsi réalisés sur une période d’un an, dans des conditions météorologiques diverses. La dispersion obtenue est de l’ordre de 20 à 25 dB, elle est nettement marquée, même à des distances relativement proches (clôture), et plus importante à 125 Hz qu’en dB(A) (voir fig. 2). En fait l’écart interquartiles est compris entre 4 et 8 dB(A) et 7 à 10 dB dans la bande de 125 Hz.

Etablissement d’un coefficient de propagation moyen

A partir des puissances acoustiques exactes des appareils, un centre acoustique équivalent a été déterminé de façon à obtenir la relation suivante:

$L = N_w - K_n \log d - 8$

dans laquelle $L$ désigne le niveau de pression (dB(A) ou bande de 125 Hz) à la distance $d$ (en m), $N_w$ le niveau de puissance totale de la source (dB(A) ou bande de 125 Hz) et $K_n$ le coefficient de propagation correspondant.

La position de ce centre acoustique équivalent n’est pas fixe, elle dépend du point de mesure considéré (du fait que les 12 transformateurs, de puissances acoustiques parfois différentes, sont vus de chaque point considéré selon des distances et des angles variables). La distance d du point de mesure au centre acoustique équivalent est ainsi ajustée pour chaque point de mesure.

La valeur moyenne du coefficient $K_n$ a été trouvée égale à 23.4 en dB(A), sans grande distinction de jour ou de nuit, et à 26.7 le jour ou 25.3 la nuit dans la bande de 125 Hz, ce qui laisse apparaître une influence nocturne sur la propagation des basses fréquences (voir fig. 3 et 4). Calculée par axe, elle ne varie que de 0.3 de jour, et de 0.5 (en dB(A)) ou 0.8 (à 125 Hz) la nuit; on peut donc tendre à conclure sur une influence très faible de la topographie et de la présence de différents boisseaux selon les trois directions de propagation en regard des effets plus importants de certains paramètres climatiques. Il est en fait possible que l’influence exercée par l’impédance du sol – étudiée par de nombreux auteurs sur des distances relativement courtes [1, 2, 3] – soit contrebalancée par les divers phénomènes physiques rencontrés sur des distances plus importantes [1, 4], tels que la réfraction ou les turbulences.

A titre de vérification, une régression entre le niveau de pression et la distance a été établie, pour chaque relevé sur un même axe de propagation, sous la forme:

$L = (N_w - R) - K_n \log d$

Les valeurs du coefficient de propagation $K_n$ ainsi obtenues fluctuent entre 10 et 35, cette fluctuation étant plus importante à 125 Hz et la nuit. Par contre, il est intéressant de noter que des puissances acoustiques équivalentes moyennes obtenues de toutes les régressions (strictement à partir des moyennes sur tous les relevés) (ainsi que les valeurs connues à la source) sont, pour la bande de 125 Hz, pratiquement égales aux puissances totales réelles (125.7 dB la nuit pour une réalité pratique de 122.3 dB).

Relations avec les paramètres climatiques

Sur le plan théorique, l’influence des différents paramètres climatiques a été revue de façon détaillée par plusieurs auteurs [4, 5]. Néanmoins, il n’existe pas de modèle sûr applicable à la propagation près du sol sur des distances supérieures à 600 m [6].

A partir des données horaires de la station météorologique installée pour le projet, un indice de gradient de température a tout d’abord été construit à partir des deux pentes relevées entre 3 et 73 m, et entre 73 et 137 m: pour cela, la première pente a été considérée, du gradient positif ou négatif, en ajoutant l’effet des inversions de température, suivant leur valeur notée à l’aide de la seconde pente. Les valeurs de cet indice se situent entre 1 et 17, avec des moyennes respectives de jour et de nuit de 4 et 13, ce qui laisse déjà présager une meilleure propagation nocturne.

Un indice de vent a été ensuite réalisé en utilisant la vitesse et la direction relevées à 73 m; ces chiffres témoignant d’une meilleure stabilité que ceux à 6 m. Le vecteur vent correspondant à l’heure moyenne des relevés a été projeté sur chacun des trois axes de propagation étudiées, en éliminant ce-pendant tous les dossiers pour lesquels la vitesse était supérieure à 25 km/h.

Enfin, à partir des tableaux classiques d’atténuation linéaire [7], nous avons bâti un indice
d'atténuation en $10^{-2}$ dB/100 m. Devant la faiblesse des corrélations obtenues, nous avons utilisé un indice plus simple, directement relié à la valeur de l'humidité relative et dimensionné de l'à 15, avec lequel les moyennes de jour et de nuit se sont bien différenciées (7 et 12 respectivement).

Résultats préliminaires relatifs aux influences climatiques

Etant donné l'importance des relevés compilés, nous avons encore procédé qu'à des analyses simples de régression entre les paramètres climatiques et les mesures acoustiques rassemblées sous la forme, décrite précédemment, soit celle des coefficients de propagation par axe et par relevé. Cependant, il apparaît déjà que les deux paramètres les plus importants sont, dans l'ordre des coefficients de corrélation, le gradient de température et l'humidité relative, ceci avec des contributions bien significatives à la propagation, surtout dans la bande de 125 Hz et en période nocturne.

Quant à l'effet du vent, que plusieurs auteurs semblent valoriser en conjonction avec les autres paramètres [8], ses corrélations sont faibles et ne semblent indiquer, sur une base statistique, que l'effet de la distribution géoclimatique habituelle du site (réduction de la propagation sur l'axe ouest par exemple). Par ailleurs, on note que les fluctuations du niveau de pression sont souvent plus importantes en présence d'un fort vent, tel que mentionné dans la littérature [4].

En fonction de ces premiers résultats, nous nous attachons actuellement à construire un modèle statistique simple qui puisse, notamment à partir du gradient de température et de l'humidité relative, refléter la plus grande variance possible des niveaux de pressions. Ceci afin de modéliser, pour une situation climatique donnée (mois de l'année, période de la nuit ou du jour), la dispersion possible des atténuations prévisibles jusqu'à 2 km.

Remerciements

Nous remercions la Société d'État Hydro-Québec pour avoir installé la station météorologique automatique et contribué financièrement au projet de recherche, de même que le ministère fédéral de l'Énergie, des Mines et des Ressources du Canada.

Références


FIGURE 1: Exemple d'enregistrement des niveaux nocturnes à 1250 m (2h) à 3h30.

FIGURE 2: Distribution des niveaux de pression.

FIGURE 3: Atténuations supplémentaires moyennes dans la bande de 125 Hz.

FIGURE 4: Distribution du coeff. de propagation.
INFLUENCE DU SOL ET DES VARIATIONS ATMOSPHERIQUES SUR LA PROPAGATION SONORE A GRANDE DISTANCE : EXPERIMENTATIONS REPETES SUR QUATRE SAISONS

M.C. Bérengier *, G.P. Oswald **, B. Mériel ***

(*) L.C.P.C. Centre de Nantes, BP 19, 44340 Bouguenais, France
(**) K.D.F., DEP. ACoustique, 1 Av. du Général de Gaulle, 92141 Clamart Cedex, France
(***) LRPC, Rue Laplace, 41000 Blois, France

La propagation acoustique en milieu extérieur est influencée par un grand nombre de facteurs : effet du sol, effets atmosphériques, diffusion, turbulence, etc. Dans le but de valider un modèle numérique de propagation atmosphérique (1) et d'investiguer une banque de données (mesures acoustiques et météorologiques), des séries d'expérimentations de propagation sonore ont été réalisées sur site au cours des quatre saisons.

DESCRIPTION DES ESSAIS

Cette campagne de mesures s'est déroulée d'octobre 1984 à juin 1985 sur un site dégagé dont le sol est plan et composite (sol barbeux, sols en culture, pises en béton). La source de simulation composée de trois haut-parleurs pneumatiques de forte puissance (> 130 dB(A) à 1 mètre) émet un bruit rose sur la bande utile (150 Hz, 3 kHz). Des enregistrements acoustiques sont effectués simultanément à deux hauteurs (2 et 6 m) en divers points de site repartis sur quatre circonférences situées à 30 m (référence), 100, 600 et 900 m de la source. Parallèlement, à partir d'un mat de 23 m de hauteur, on relève les profils verticaux de quatre paramètres atmosphériques : température, vitesse et direction du vent, humidité relative. Pour tenir compte de l'absorption supplémentaire due aux différentes natures de sol, des mesures d'impédance caractéristique sont effectuées à partir d'une méthode de calcul utilisant un signal transitoire (2).

Les quatre campagnes d'essais réparties sur les quatre saisons ont permis d'accéder à 30 situations météorologiques différentes. Les mesures d'impédance de sol ont été effectuées à chaque campagne pour une trentaine de points, ce qui constitue un catalogue d'environ 120 types de sol. Enfin plus de 800 enregistrements acoustiques de 45 mm chacun ont été collectés au cours de cette même période.

RESULTATS DE MESURES

Impédance de sol

Affin d'accéder à l'impédance caractéristique Z, on mesure tout d'abord le module d'atténuation sonore entre deux points, sous incidence rassante, puis on utilise une procédure de calage afin d'optimiser les cinq coefficients d'un modèle paramétrique qui se présente comme suit:

\[
\cRe(Z_0) = \frac{1}{a_1} \left( \frac{a_2}{a_2} \right) \quad \cIm(Z_0) = \frac{1}{a_4} \left( \frac{a_2}{a_2} \right)
\]

Au cours de cette opération, l'atténuation théorique est calculée à partir d'un modèle de propagation de Thomsson considérant le sol comme un milieu à réaction localisée (3). Aprés calcul du déphasage excédentaire, Z est obtenu en appliquant une méthode décrite dans (4). Les résultats portés sur les figures 1 et 2 sont représentatifs d'un sol barbeux.

En reproduisant l'opération à tous les points du site au cours des quatre périodes, on a obtenu les comparaisons suivantes :

Tableau 1 : Evolution de l'impédance Z et de l'absorption à d'un même sol au cours des quatre saisons

<table>
<thead>
<tr>
<th></th>
<th>250 Hz</th>
<th>510 Hz</th>
<th>750 Hz</th>
<th>1010 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>(06/85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(06/85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10/84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(04/85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(12/84)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tableau 2 : Classification de plusieurs sols lors d'une même campagne de mesure (06/85)

1=sol couple, herbe épaisse
2=sol herb épaisse
3=sol herbe haute
4=sol dur

<table>
<thead>
<tr>
<th></th>
<th>250 Hz</th>
<th>510 Hz</th>
<th>750 Hz</th>
<th>1010 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ces résultats permettent de retrouver que de façon générale, l'absorption est d'autant plus importante que le support est souple et le couvert végétal dense.

Corrélation bruit-effets météorologiques

A chaque série de mesures acoustiques est associé un état météorologique complet. Les variations de niveaux de bruit relevées lors d'une même série sur les différents points de mesure (cf. figure 3) s'expliquent principalement par l'effet du vent (force et direction) (cf. figure 4). On observe notamment des écarts importants (de l'ordre de 20 dB) suivant que l'on soit en vent portant ou contraire, à sol quasiement identique.

COMPARAISONS MESSURE-CALCUL

Cette opération a été effectuée pour le 1/3 d'octave 500 Hz et pour des points de mesure situés à 2 m de hauteur et à 600 m de la source, celle-ci étant placée à une hauteur de 10 m. En octobre 84, la vitesse du vent était de 5 m/s, sa direction de 190° par rapport au nord et la puissance de la source de 85 dBA à 30 m. En avril 85, la vitesse du vent était de 7.5 m/s, sa direction de 45° par rapport au nord et la puissance de la source de l'ordre de 100 dB à 30 m.

Tableau 3 : Comparaison mesure-calcul en dB pour tous les points situés à 600 m pour les campagnes d'octobre 1984 et avril 1985 (1/3 d'octave 500 Hz)

<table>
<thead>
<tr>
<th>Points</th>
<th>Octobre 1984</th>
<th>Avril 1985</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Messure</td>
<td>Calcul</td>
</tr>
<tr>
<td>B</td>
<td>57.5</td>
<td>57.7</td>
</tr>
<tr>
<td>C</td>
<td>43.7</td>
<td>41.7</td>
</tr>
<tr>
<td>D</td>
<td>34.8</td>
<td>35.4</td>
</tr>
<tr>
<td>E</td>
<td>57.2</td>
<td>57.3</td>
</tr>
<tr>
<td>F</td>
<td>58.6</td>
<td>59.4</td>
</tr>
<tr>
<td>G</td>
<td>57.7</td>
<td>57.8</td>
</tr>
</tbody>
</table>
Des travaux antérieurs (1) ont montré que la seule prise en compte de l'effet de sol conduisait à un résultat de calcul assez éloigné de la réalité expérimentale, notamment lorsque le sol était contrai-
re. Le tableau comparatif ci-dessus montre que la combinaison des deux effets sol et météorologique, permet de rattraper les écarts importants précédemment mis en évidence. Malgré un accord dans l'ensemble très satisfaisant, il est possible de relever quelques écarts substantiels (\(5 \text{ à } 6 \text{ dB}\)) sur certains points, ceci s'expliquant soit par une mesure erronée de l'impédance (sol labouré ou cultivé peu assimilable à une réaction localisée), soit par une définition trop grossière des profils atmosphériques.

CONCLUSION ET PERSPECTIVES

Ces essais expérimentaux qui ont d'une part permis de valider un modèle numérique de calcul de propagation en milieu extérieur permettront d'autre part de rechercher des corrélations statistiques entre le bruit et les effets météorologiques en fonction des sols rencontrés et d'établir un catalogue de sol en vue d'utilisation extérieure.

BIBLIOGRAPHIE

(1) C.P. Oswald, La Revue d'Acoustique, Vol 72, 1985, pp 128-140
(3) S.I. Thomas, JASA, Vol 59, n° 4, 1976, pp 780-785

Figure 1 : comparaison des modules d'atténuation
- expérimenté ; - - - - : modèle

Figure 2 : Variation de l'impédance en fonction de la fréquence
(a) : partie réelle ; (b) : partie imaginaire
- - - - : Variation obtenue par transformation de Hilbert

Figure 3 : courbes d'atténuation mesurée à 600 m autour de la source
1/3 d'octave 500 Hz ; \(h_s = 10 \text{ m} ; h_f = 6 \text{ m}\)

Figure 4 : Force et direction du vent.
L'expérience a montré que l'hypothèse de déphasage minimal n'est pas restrictive et qu'elle est vérifiée sans difficultés tant que l'hypothèse de réaction localisée est valable.

On constate toutefois que le domaine d'intégration de cette relation s'étend sur l'ensemble du domaine fréquentiel de zéro à l'infini alors que le modèle n'est connu expérimentalement que dans la bande \([0, \delta]\). La contribution nécessaire de la bande \([\delta, \infty]\) est fournie par un prolongement spectral du modèle dans cette bande fréquentiellement déterminé de la façon suivante:

- on utilise un modèle analytique décrivant le rapport des spectres des impulsions
- par un calage au sens des moindres carrés, on ajuste ce modèle sur le module expérimental dans la bande \([0, \delta]\)
- on assimile le comportement du module du modèle dans la bande \([\delta, \infty]\) à celui du modèle expérimental

Modèle analytique du rapport des spectres

D'après Thomasson, le champ acoustique \(p\) émis par une source ponctuelle \(S\) de coordonnées \((x_S, y_S, z_S)\) dans un demi espace limité par un dioptr plan avec un comportement à réaction localisée, caractérisé par son impédance spécifique \(Z\) ou son adhérence \(v = 1/z\), peut s'écrire ; en un point \(R\) (\(x_R, y_R, z_R\))

\[ p = p_D + p_R + p_P + p_S \]

avec \(p_D = \exp(jk_0 R_1)/4\pi R_1\) \(p_R = \exp(jk_0 R_2)/4\pi R_2\)

\[ p_R = \frac{k_0}{w_0} \int_0^{\infty} \frac{e^{-ct}}{w_0} W(t) dt \]

\[ p_S = \frac{1}{\sqrt{2 \pi}} \int_{-\infty}^{\infty} e^{j\omega t} \overline{W(t)} dt \]

ou

\[ W(t) = (A^2 + B^2)^{\frac{1}{2}} \]

\[ A = (j k_0 R_2 (y_t - i)^1/2 \gamma) \]

\[ B = (j k_0 R_2 (1 - i y_t)^1/2 \gamma) \]

\[ \gamma = \cos \theta_r - (1 - \omega^2)^{1/2} \sin \theta_0 \]

\[ \gamma = \cos \theta_r - (1 - \omega^2)^{1/2} \sin \theta_0 \]

\[ C_0 = \int_0^1 + \int_{\theta_0}^{\pi/2} \pi \delta \kappa \alpha \kappa / 2 \]

\[ R_2 = (r_2^2 + z_2^2)^{1/2} \]

\[ R_1 = (r_1^2 + z_1^2)^{1/2} \]

Le modèle analytique du rapport des spectres est donné par le rapport du champ \(p_2\) au point \((x_2, y_2, z_2)\) et du champ \(p_1\) au point \((x_1, y_1, z_1)\) et \(z_2\) sont les hauteurs des deux microphones, théoriquement nulles, en pratique elles sont de l'ordre de quelques centimètres.

\[ |H(\omega)| = \frac{P_2(x_2, z_2, y_2, \psi, \omega)}{P_1(x_1, z_1, y_1, \psi, \omega)} \]

Pour réaliser le calage du module théorique sur le module expérimental dans la bande d'analyse \([0, \delta]\), nous ajustons les paramètres d'un modèle d'impédance.

Modèle d'impédance

Plusieurs définitions de modèles ont été testées. Nous avons constaté que celle qui permettrait de simuler le plus fidèlement l'allure du module expérimental dans toute la bande d'analyse est de la forme :
\[ Z(f) = 1 + \frac{a(f)}{B} + j \cdot \frac{c(f)}{B} \]  

(3)

\( f \) est la fréquence, \( a, B, \gamma, \delta, c \) sont les paramètres du modèle dont les valeurs sont déterminées par le calage des modules.

Ce modèle a une forme analogue à celui proposé par Delany et Bazley mais pour les matériaux fibreux, il possède cinq degrés de liberté et permet un ajustement du module du rapport des champs acoustiques autant dans la partie basse fréquence que dans la partie haute fréquence de la bande de mesure.

RÉSULTATS EXPERIMENTAUX

Ce processus de détermination de l’impédance a été appliqué à plusieurs natures de sol. Nous présentons ici les résultats obtenus pour le même sol mais à des époques et conditions météorologiques différentes. Ils concernent un sol végétal recouvert d’herbe situé à une centaine de km à l’ouest de Paris : la première série de mesure a été faite au début de l’hiver au mois de décembre sous une température ambiante de 0°C, la seconde au cours du mois d’avril du printemps suivant sous une température de 11°C. Les valeurs respectives des paramètres du modèle d’impédance déterminées par le calage du modèle sur le module expérimental sont : pour la série n° 1 \( a = 5.2, \beta = 258, \gamma = -2.7, \delta = 8.6, c = -0.55, \) et pour la série n° 2 \( a = 1.81, \beta = 304, \gamma = -1.05, \delta = 6.2, c = -0.30 \)

Les figures ci-jointes montrent les parties réelle et imaginaire de l’impédance des deux séries : les courbes en trait discontinu sont relatives au modèle d’impédance, celles en trait continu sont les valeurs d’impédance obtenues par inversion de la formule (2) ; on constate ainsi le bon accord entre l’allure des courbes données par le modèle (3) et les courbes obtenues par le calcul à partir des valeurs expérimentales.

CONCLUSION

Le modèle d’impédance choisi paraît donc être une approche assez fiable de l’impédance d’un sol à réaction localisée. Il ne semble pas utile de déterminer par le calcul les parties réelle et imaginaire de l’impédance à toutes les fréquences, la détermination des cinq paramètres du modèle d’impédance suffit pour avoir une estimation des composantes de l’impédance dans toute la bande de fréquences intéressante.

REFERENCES

Thomasson S.I. Reflection of wave from a point source by an impedance boundary JASA 1976, volume 59, n° 4, p. 780-785
Leganey V. Application de la transformée de Hilbert discrète à l’évaluation de coefficients de réflexion acoustiques, thèse 3ème cycle, Université du Maine, le Mans, France, 1981
STANDARDIZED PROCEDURE FOR RATING INDUSTRIAL NOISE

Lucia N. Talbo

Instituto Nacional de Tecnología Industrial, Dto. de Física, División Acústica, C.C. 157, 1650 San Martín, Prov. de Bs. As., R.A., Argentina.

INTRODUCTION

Noise from industrial installations affecting people in mixed residential areas is a matter of permanent conflict. In our cities, the lack of an appropriate urban development, in order to separate areas according to the main activities, allows the close proximity between factories and dwellings. In this context, the direct interconnection between factories and dwellings through a solid wall or dividing wall is usual to be found, thus allowing the transmission of both direct and structureborne sound.

In spite of the great difficulty in evaluating each individual response, the need of an objective procedure is rather obvious. New trends in standardization are mainly oriented towards the measurement of exterior "A" levels for rating whole areas in order to establish noise limits [2]. A German standard [5] considers noise limits inside dwellings, for day or night time.

Our problem is to know to which degree an identified sound source is influencing people in a specified location within a building. Intrusive noise has to be determined within a building, since other levels could perhaps add no information about the real problem. The extent to which noise enters a house is difficult to predict, following simple considerations, since depends on the combination of a multiplicity of sound paths, both airborne or structureborne. In many cases, exterior noise measurements should probably give a level within acceptable limits for the whole area. For example, in a mixed residential area, outside level was 54 dB(A), within the acceptable limit of 55 dB(A). Inside level was 35 dB(A), producing a clear sensation of discomfort. The background level was only 26 dB(A), in the back rooms.

STANDARDIZED PROCEDURE

The procedure has to evaluate the degree of intrusiveness of the noise from a specified source, in a certain location. Accordingly, the noise emitted from a specified source, its characteristics, both temporal and spectral and the background or preexisting noise level, must be known.

Some years ago, a procedure has been standardized [4], based on BS 4142 [3] and following the general assumptions of [1]. Basically, the noise to be rated, L, is measured in dB(A) and "slow" response. For variable noise, L eq is determined during a representative period and then, corrected by its tonal character and impulsiveness, thus determining the rating level, L r. The background level, L b, defined as the minimum dB(A) level (statistically as L 90), the noise source off and without the inclusion of occasional extraneous sound, is measured in the same location and conditions as L. Besides, a calculation is made for estimating the typical or calculated background level, for the specified situation, L c, from a basic level, plus correction factors such as, type of area, time of the day, etc., determined from experience. An excess margin, M, is then defined as the difference between L r and L b or L c, whichever the lowest. If the margin exceeds a specified absolute value the noise is rated as annoying, for the considered situation.

AVAILABLE DATA

During several years a lot of data has been collected, by the systematic application of the above mentioned procedure [4]. About 400 hundred dwellings have been surveyed, thus gathering information about noise levels and opinions about noise characteristics (tonal character, discontinuities or impulsiveness), conducted by well trained people. Almost all considered cases, had been a matter of conflict. Residents' opinions about noise and the degree of annoyance were gathered and compared to standardized ratings.

EXISTING BACKGROUND LEVELS

An analysis was conducted in order to classify background levels, according to some characteristics of each specific situation, that could allow the fixation of limiting values for interior spaces [5]. This may have the advantage of simplifying the measurement procedure and consequently solve those cases, where the direct measurement of L b resulted impractical. Besides, the prediction of L b should prevent its gradual increase due to abnormal situations in a certain location. As example, the distribution of measured L b noise for a representative daytime period is shown in Figure 1;

<table>
<thead>
<tr>
<th>L B</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>173</td>
</tr>
<tr>
<td>51</td>
<td>5</td>
</tr>
<tr>
<td>58</td>
<td>2</td>
</tr>
<tr>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>66</td>
<td>1</td>
</tr>
<tr>
<td>70</td>
<td>1</td>
</tr>
<tr>
<td>73</td>
<td>1</td>
</tr>
<tr>
<td>76</td>
<td>1</td>
</tr>
<tr>
<td>79</td>
<td>1</td>
</tr>
<tr>
<td>82</td>
<td>1</td>
</tr>
<tr>
<td>85</td>
<td>1</td>
</tr>
</tbody>
</table>

(*) Lb = 45 dB
sd = 8 dB
n = 173

(0) Lb = 51 dB
sd = 5 dB
n = 173

25 30 35 40 45 50 55 60 dB
where:

(*) Daytime, Area: Predominantly urban residential, with some light industry or main roads. Interior rooms oriented to street, distant traffic.
(X) Same conditions, near traffic.
(O) Same conditions, but outer spaces of dwellings, not facing to street, both near and distant traffic.

In an attempt to estimate background levels from known variables, that can be easily determined in each case, a linear model was adjusted, considering characteristic factors of area, time of day and orientation within the building. The linear regression between estimated and measured Leq was established for 420 cases, being the prediction area at 95% confidence level, within ± 8 dB.

DISTRIBUTION OF MARGINS

In Figure 2, the percentage of margins distribution is shown. They represent the result of measured cases, where complaints or legal actions were conducted by residents.

```
+---------------------------+---------------------------+
| % | "A" differences |
+---------------------------+---------------------------+
| 10 |                            |
| 20 |                            |
| 30 |                            |
| 40 |                            |
| 50 |                            |
+---------------------------+---------------------------+
 Figure 2: Distribution of margins
```

The average is about 12 dB, with ad = 6 dB. About 66% of the observed cases correspond to margins greater than 10 dB. A clear sensation of discomfort was present. Approximately 22% of the cases laid between 5 to 10 dB. This indicates a transition zone where it is possible to expect complaints. Below 5 dB, almost 15% of the observations have margins ranging from 0 dB to 3 dB, that contrarily as could be expected, most of them correspond to highly annoying situations, where a clear sensation of discomfort is noticeable for both residents and trained people.

INFLUENCE OF SPECTRAL SHAPE

Many authors [6], [7], studied the influence of noise spectral content, specially in those cases, where a high concentration of low frequency energy is present. When airborne noise is entering from outside, the acoustical behaviour of partition walls, together with the modal response of the rooms, tend to emphasize the unbalancing of spectra, increasing low frequencies with respect to the middle and high range.

Some rating procedures were studied during last years. Most recently [8], a classification is proposed following a similar procedure as for NR curves, but with a modified set of curves, LPFN. A 1/3 octave analysis of noise, is needed. Normally, for practical purposes, it is not able to know the 1/3 octave spectra with conventional SLMs. As a first approximation to detect the problem, linear SPL margins, instead of "A" margins, were determined. Accordingly, the low frequency increase, with respect to background, masked by dB(A), was evident.

For those most frequent cases, where the introducing noise presents an increment in the middle and high frequency range, "A" differences could be more sensible than LSPLs, specially when the low frequency content of background noise is high. The method proposed can be rated to take either "A" differences or LSPL, whichever is the largest.

CONCLUSIONS

The need of measuring background levels is obvious, due to the large variation in existing levels and in its prediction for a given situation.

The determination of margins from LSPL differences, allows to explain the strong annoying sensation, in cases where "A" margins ranged from 0 dB to 3 dB. On the other hand, when high "A" margins are found, LSPL margins could mask the increase in level, due to the high low frequency content of background noise. Making a distribution of margins following this analysis, from 10% of unexplained cases, only 2% remained in this level. In those cases, the actual situation was informed as "not noticeable" by technicians. Nevertheless, residents assured to feel annoyed by the noise. More extensive and deep studies are probably needed to answer this question.

REFERENCES

Abdel-Alla, A. O. A-3-1, B9-2
Abel, N., A-5-3, B10-3, B10-5
Abraham, P. W. C5-7
Abramov, O.V. G6-4, G6-5
Abrhy, C. A3-8, A6-1
Ackenhusen, J.C. M1-6
Adler, L. G3-4, I3-7
Agnan, J.-L. M2-4
Agulló, J. R6-5
Akgül, M. A2-6
Akulichev, V.A. T4-5
Alberti, F.W. B10-3
Alippi, A. D5-7
Allard, J.F. I5-4, M3-2
Allen, C.H. B9-5
Allie, M.C. G5-2
Almen, M. E11-2
Amara, M. J2-4
Anderson, J.S. E10-2
Ando, Yoichi Plenary 2, E14-14, E11-3
Ando, Yoshinori K1-5
Andrew, C. C9-4
Anema, P.C. B4-1
Angeles, M. D6-9
Apfel, R.E. 24-3
Arbey, H. B5-4, C3-3
Archibald, J. A6-3
Aram, K. C5-1, C5-3
Assayma, K. M1-4
Asselman, M. C4-5, C4-8, C10-7
Atherley, G.R.C. E2-4, B7-1
Attenborough, K. B5-1
Auletta, N. E12-5
Avallet, C. D3-7
Avan, P. B6-2
Avery, P. A6-3
Awbrey, F.T. J2-7
Ayres, R.D. K1-6
Ayres, V.M. T1-5
Azais, C. B4-3
Bae, J.-R. G1-1
Bagnoli, S. A5-1
Bail, N. I3-1
Bakar, C.A. C9-7
Bakken, M.J. B2-4
Bardawe, O. D1-7
Bansal, A.S. D5-1
Barby, J.L. E12-3, E12-8
Barkeikh, M. K1-2
Barlow, M. E6-3
Baruch, C. B2-8
Baruh, M. D4-4
Beach, S.H. B8-3
Bech, S. L2-7
Bedard, A.L. J2-5
Bequeux, B. C8-9, C8-10
Behar, A. B10-1
Behler, C. E9-3
Bclenuz, G. K1-4
Bell, J. B8-6
Belteroibaž, F. C4-7, K3-9
Benade, A.R. K6-3
Benatot, D.R. J2-1
Bennett, D. M4-6
Bennighof, J.K. O4-5
Benoit, C. A6-1
Bento Coelho, J.L. M3-10
Benwell, D.A. C9-6, F1-5
Bergman, W. C2-3
Berenger, M.C. C10-8, C10-9
Berger, K.H. B9-1, B9-5
Berkhout, H. A. E4-12, E4-4
Bertason, A.C.O. E6-1
Berry, A. E12-2
Berzolet, M. O5-7
Bickley, C. A6-4
Billaud, J. E12-7
Bilien, P.A. E6-5
Birou, D. H2-10
Bisaro, P. B6-2
Bite, M. D1-6
Björkä, L. M3-8
Blackstock, D.T. Plenary 3, E2-1
Blaise, A. D3-3
Blanc-Benou, Ph. J1-6
Blauret, J. B1-6, B2-9
Bly, S.H.P. F1-5
Bodemann, O.F. L2-4
Boe, J.L. A4-6
Botella, P.E. F2-8
Boisson, C. E8-6
Boistad, E.H. C9-4
Bolt, J.S. E1-5
Bonti, H.G. B10-2
Bonne, M.M. C4-4
Booth, J.C. B1-2
Bottros, R. E9-2
Bottre, M.-C. B2-8
Bourdier, K. T5-4, M3-2
Boutillon, X. K2-6
Boutin, J. F2-8
Bowes, L. A3-9
Brackett, S.E. C2-3
Bradley, J.S. E10-5
Braunma, A.J. F2-4, F2-5, F2-6
Breazeale, M.A. G4-5
Breitschwerdt, K.G. C2-6
Brill, D.W. N1-8
Brinkmann, K. M9-2
Brooke, G.R. H2-6
Brown, S.M. D8-5
Brown, W.S. B1-3
Brüel, P.V. E10-3
Bruneau, M. M5-7, M3-5
Bruneau, M. K3-5
Bacherr, O. G2-7
Budisavljević, B.B. E8-2
Bull, M.K. J3-4
Bunke, S.S. J2-6
Bunn, A.E. B8-2
Burns, E.M. B7-5
Busch-Vishniac, I. L1-2
Bus, S. B11-3

- C -

Cabot, R.C. M4-5
Cachard, C. G1-6
Cafaze, M.K. K12-3, K12-8
Caliokan, M. C2-9
Candel, S.M. J1-7
Cantrell, J.H. O3-1
Fujinawa, K.        E7-5
Fukuda, K.          L2-8
Fukudome, K.        B6-3, B9-4
Fukushima, M.       C3-1, C3-2
Funai, H.           B6-1
Furui, S.           A2-6

- G -
Gagliardi, L.       E1-4
Gaja Diras, K.      C4-7
Gallego Juárez, J.A. G4-2, G6-2
Gamba, R.           G7-2
Gan, W.S.           G7-6
Garntner, G.R.      C10-3
Gaspar, K.          C2-3
Gatehouse, R.W.     B2-7
Gauer, J.L.         D2-1
Gauwaard, G.C.      H1-2, H1-5, H1-8
Gawne, K.           E12-9
Gemuir, K.          B6-8
Gerges, S.N.Y.      E7-7
Gerhardt-Multhaupt, R. G7-5
Getty, L.           B1-8
Geveling, S.M.      D2-3
Gewurtz, S.         C7-1
Gharrabegian, A.    E10-10
Ghosh, S.           G1-4
Giani, S.           B2-8
Glani, A.           A3-2
Glerlich, R.W.      B4-8
Giguere, C.         B10-5
Gif, C.             G6-9, M4-8
Gleizes, A.         K3-9
Gleizes, G.         G1-6
Glerard, L.         C11-8
Cjestland, T.       C4-1
Glegg, S.A.L.       J3-6
Gong, N.B.          C6-5
Goodman, D.J.       D1-1
Goodman, R.E.       Finary 1
Gordon, D.F.        H3-6
Goubard, R.         E9-2
Goudailier, J.-P.   A3-7
Grédiac, C.         C7-6
Greaves, T.J.       J3-4
Green, K.K.         E1-5
Gren, L.H.          H1-6, H1-7
Greene, G.E.        J2-5
Greiner, R.A.       C5-2
Griefahn, B.        C4-3
Griffin, M.J.       Finary 5
Grönig, H.          H3-3
Gudmundsson, S.     E8-5
Gueske, R.W.        B8-2
Guérin, B.          A3-8, D3-3
Guerrero, M.        K5-4
Guilhot, J.P.       B4-3, C8-6
Guimel, F.          C8-4, D3-3
Gulyaev, Yu.V.      Finary 7
Gumbringer, M.D.    K3-3
Gut, S.             G2-6
Guy, R.W.           E1-2
Guyader, J.-L.      E1-3, E1-4, E8-6

- H -
Habault, D.         C8-3, M3-4
Hashiya, N.         H1-3
Haddad, S.D.        C2-5
Haering, H.-U.     C2-7
Hama, H.            R4-7, B6-1, C5-1
Hamann, D.          I3-9
Hamlet, M.F.        C6-2, I2-4
Hamon, L.           E7-9
Hansen, U.J.        E2-8
Hansrote, R.        E5-2
Harmegnies, B.      A1-2, A1-4
Harris, J.C.        G5-6
Harvenberg, C.E.    A4-2
Hasebe, M.          J1-3
Hatakeyama, K.      C1-6, C4-6, E2-8, E2-9, E2-10
Hathaway, K.B.      L1-3
Hayek, S.I.         D3-9
Heckl, M.           D2-6
Hellbrück, J.       B4-4
Hellman, R.P.       B11-5
Henderson, D.       B3-2
Hendrickson, T.J.   K2-4
Henriet, C.         E7-9
Henrio, J.G.        C8-2, C8-5
Herbert, K.         J3-3
Herevan, M.         C7-1
Herzog, P.          I3-5
Hifci, R.           B3-4, B5-3, B5-7, B5-8, B11-7, B11-8
Hillman, R.B.       C1-5, C7-3
Hind, J.E.          B6-5
Hirst, D.J.         A2-7
Hodgey, J.P.        F2-7
Hodgson, H.         B12-1
Hohm, D.            I3-3
Højbjerg, K.        E7-8
Holding, J.M.       M4-2
Holmsen, E.B.       A3-10
Holmes, M.H.        B7-4, B8-6
Holt, J.A.          M2-2
Hong, R.K.          M3-8
Horie, G.           C1-4
Hornowsk, T.        G7-7
Hothwareall, D.C.   J1-2
Huttebard, C.       G5-8
Houtgast, T.        E6-13
Houtemans, A.J.M.   B2-1
Howell, W.E.        M1-3
Hriblick, M.P.      L3-4
Hu, J.              L2-2
Hubbard, J.K., Jr.  D4-1
Hühbner, C.         D2-7
Hughes, D.O.        C2-6
Hunt, M.J.          A4-2
Huston, R.D.        I1-10
Hutchins, C.M.      K3-4, K3-5
Hutchins, D.A.      G4-3, L2-2
Hutchins, M.A.      K3-4

- I -
Ibbas, G.           A6-10
Ichinomiya, O.      D2-4
Idec, M.            F1-4
Idogawa, T.         K4-2
Iida, K.            E11-3
Imai, A.            E3-2
Imamura, T.         G5-1
Ingram, D.          A2-3
Inman, C.           E2-1
Intriario, J.       J2-5
Iseki, T.           J1-5
<table>
<thead>
<tr>
<th>Last Name</th>
<th>Initials</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ishibashi, M.</td>
<td>K=2</td>
<td></td>
</tr>
<tr>
<td>Ito, M.R.</td>
<td>A1-7</td>
<td></td>
</tr>
<tr>
<td>Iwamiya, S.</td>
<td>B2-2</td>
<td></td>
</tr>
<tr>
<td>Iwashige, N.</td>
<td>E7-5</td>
<td></td>
</tr>
<tr>
<td>Izbički, J.L.</td>
<td>G1-3, I1-3</td>
<td></td>
</tr>
<tr>
<td>Jackson, R.A.</td>
<td>B10-1</td>
<td></td>
</tr>
<tr>
<td>Jacob, A.D.</td>
<td>G7-8</td>
<td></td>
</tr>
<tr>
<td>Jacobsen, T.</td>
<td>E10-1, E10-2</td>
<td></td>
</tr>
<tr>
<td>Jacques, P.</td>
<td>C4-8</td>
<td></td>
</tr>
<tr>
<td>Jamieson, D.G.</td>
<td>A2-5, A5-2</td>
<td></td>
</tr>
<tr>
<td>Jänkäinen, M.G.</td>
<td>E1-1</td>
<td></td>
</tr>
<tr>
<td>Jančkovič, S.</td>
<td>J2-3</td>
<td></td>
</tr>
<tr>
<td>Jarosz, B.J.</td>
<td>C4-1</td>
<td></td>
</tr>
<tr>
<td>Jensen, F.S.</td>
<td>H2-5</td>
<td></td>
</tr>
<tr>
<td>Johnson, T.</td>
<td>A2-5</td>
<td></td>
</tr>
<tr>
<td>Johnson, W.T.</td>
<td>J3-4</td>
<td></td>
</tr>
<tr>
<td>Jonasson, N.G.</td>
<td>E2-2</td>
<td></td>
</tr>
<tr>
<td>Jones, K.</td>
<td>B7-5</td>
<td></td>
</tr>
<tr>
<td>Jossen, F.</td>
<td>C7-2</td>
<td></td>
</tr>
<tr>
<td>Joufflet, P.</td>
<td>J1-7</td>
<td></td>
</tr>
<tr>
<td>Jouhanneau, J.</td>
<td>E6-4</td>
<td></td>
</tr>
<tr>
<td>Julien, J.P.</td>
<td>E4-9, E5-1, E6-4</td>
<td></td>
</tr>
<tr>
<td>Jungman, A.</td>
<td>G3-2</td>
<td></td>
</tr>
<tr>
<td>Juviš, D.</td>
<td>J1-6</td>
<td></td>
</tr>
<tr>
<td>Kahraman, A.</td>
<td>C2-9</td>
<td></td>
</tr>
<tr>
<td>Kallash</td>
<td>G2-9</td>
<td></td>
</tr>
<tr>
<td>Kakusho, H.</td>
<td>A1-6, A3-3</td>
<td></td>
</tr>
<tr>
<td>Kalb, J.R.</td>
<td>C10-3</td>
<td></td>
</tr>
<tr>
<td>Kani, H.</td>
<td>G5-1, G5-5</td>
<td></td>
</tr>
<tr>
<td>Kaneko, T.</td>
<td>A4-5</td>
<td></td>
</tr>
<tr>
<td>Kaneyasu, K.</td>
<td>J1-5</td>
<td></td>
</tr>
<tr>
<td>Kanno, S.</td>
<td>B1-1</td>
<td></td>
</tr>
<tr>
<td>Kasalainen, M.A.</td>
<td>B2-3</td>
<td></td>
</tr>
<tr>
<td>Kato, Y.</td>
<td>E2-10</td>
<td></td>
</tr>
<tr>
<td>Kitzmiller, D.</td>
<td>I3-9</td>
<td></td>
</tr>
<tr>
<td>Keane, D.</td>
<td>K5-6</td>
<td></td>
</tr>
<tr>
<td>Kearns, J.A.</td>
<td>J2-1</td>
<td></td>
</tr>
<tr>
<td>Kedrinskii, V.K.</td>
<td>I4-8</td>
<td></td>
</tr>
<tr>
<td>Keller, Y.</td>
<td>E8-3</td>
<td></td>
</tr>
<tr>
<td>Kemp, D.T.</td>
<td>B7-1</td>
<td></td>
</tr>
<tr>
<td>Kergomard, J.</td>
<td>K4-6</td>
<td></td>
</tr>
<tr>
<td>Kerry, G.</td>
<td>E1-1</td>
<td></td>
</tr>
<tr>
<td>Kersulis, A.</td>
<td>I2-5</td>
<td></td>
</tr>
<tr>
<td>Kaswick, P.R.</td>
<td>D6-4</td>
<td></td>
</tr>
<tr>
<td>Khanna, S.M.</td>
<td>B6-4</td>
<td></td>
</tr>
<tr>
<td>Khemrajjev, B.</td>
<td>G2-5</td>
<td></td>
</tr>
<tr>
<td>Kido, K.</td>
<td>G5-1, G5-5, L2-5, M2-9</td>
<td></td>
</tr>
<tr>
<td>Kihara, S.</td>
<td>H1-4</td>
<td></td>
</tr>
<tr>
<td>Kimura, M.</td>
<td>H1-1</td>
<td></td>
</tr>
<tr>
<td>Kimura, S.</td>
<td>E1-4, E11-5</td>
<td></td>
</tr>
<tr>
<td>Kinoshita, M.</td>
<td>J1-1</td>
<td></td>
</tr>
<tr>
<td>Kitamura, O.</td>
<td>B2-6</td>
<td></td>
</tr>
<tr>
<td>Kleiner, M.</td>
<td>E6-1</td>
<td></td>
</tr>
<tr>
<td>Klepper, D.L.</td>
<td>E3-2</td>
<td></td>
</tr>
<tr>
<td>Klow, T.</td>
<td>B6-6</td>
<td></td>
</tr>
<tr>
<td>Kobayashi, T.</td>
<td>A3-3, L2-1</td>
<td></td>
</tr>
<tr>
<td>Kolevlev, N.P.</td>
<td>G6-3</td>
<td></td>
</tr>
<tr>
<td>Kolevlev, Yu.A.</td>
<td>I3-10</td>
<td></td>
</tr>
<tr>
<td>Kohler, K.J.</td>
<td>A6-6</td>
<td></td>
</tr>
<tr>
<td>Komatsu, K.</td>
<td>J3-3</td>
<td></td>
</tr>
<tr>
<td>Kondo, M.</td>
<td>K3-8</td>
<td></td>
</tr>
<tr>
<td>Konishi, M.</td>
<td>E7-2</td>
<td></td>
</tr>
<tr>
<td>Konnert, J.H.</td>
<td>E4-6</td>
<td></td>
</tr>
<tr>
<td>Konno, M.</td>
<td>D1-1, M4-7</td>
<td></td>
</tr>
<tr>
<td>Kono, S.</td>
<td>B1-10</td>
<td></td>
</tr>
<tr>
<td>Konopčynski, G.</td>
<td>A6-5</td>
<td></td>
</tr>
<tr>
<td>Konovalov, N.T.</td>
<td>G6-5</td>
<td></td>
</tr>
<tr>
<td>Kor, S.K.</td>
<td>G2-10</td>
<td></td>
</tr>
<tr>
<td>Kormoch, N.S.</td>
<td>D1-9</td>
<td></td>
</tr>
<tr>
<td>Kortrijk, E.L.</td>
<td>K2-4</td>
<td></td>
</tr>
<tr>
<td>Kowalewski, J.J.</td>
<td>C3-6</td>
<td></td>
</tr>
<tr>
<td>Koyasu, M.</td>
<td>G6-7</td>
<td></td>
</tr>
<tr>
<td>Kubota, H.</td>
<td>K2-1, K3-8</td>
<td></td>
</tr>
<tr>
<td>Kuno, K.</td>
<td>C4-2</td>
<td></td>
</tr>
<tr>
<td>Kuro, N.</td>
<td>A5-3, B10-3, B10-5,</td>
<td></td>
</tr>
<tr>
<td>Kubruff, N.</td>
<td>E4-2</td>
<td></td>
</tr>
<tr>
<td>Kuwahara, S.</td>
<td>E2-8</td>
<td></td>
</tr>
<tr>
<td>Kuwano, S.</td>
<td>B3-5, C1-1</td>
<td></td>
</tr>
<tr>
<td>Laake, A.W.</td>
<td>I3-3</td>
<td></td>
</tr>
<tr>
<td>Łobovski, M.</td>
<td>G7-7</td>
<td></td>
</tr>
<tr>
<td>Lalande, N.M.</td>
<td>G3-5</td>
<td></td>
</tr>
<tr>
<td>Lardon, F.</td>
<td>K4-6</td>
<td></td>
</tr>
<tr>
<td>Lalonde, M.</td>
<td>B11-7</td>
<td></td>
</tr>
<tr>
<td>Lamb, M.</td>
<td>K1-4</td>
<td></td>
</tr>
<tr>
<td>Lambert, J.</td>
<td>B5-7</td>
<td></td>
</tr>
<tr>
<td>Landecy, A.</td>
<td>A1-2, A1-4</td>
<td></td>
</tr>
<tr>
<td>Lang, W.W.</td>
<td>C2-1</td>
<td></td>
</tr>
<tr>
<td>Langbein, C.</td>
<td>E3-7</td>
<td></td>
</tr>
<tr>
<td>Laperre, J.</td>
<td>G3-3</td>
<td></td>
</tr>
<tr>
<td>LaRoche, C.</td>
<td>B3-4</td>
<td></td>
</tr>
<tr>
<td>Latteborgh, W.</td>
<td>I4-7</td>
<td></td>
</tr>
<tr>
<td>Lawrence, A.</td>
<td>E2-3</td>
<td></td>
</tr>
<tr>
<td>Leblond, H.</td>
<td>I3-5</td>
<td></td>
</tr>
<tr>
<td>Lecomte, C.</td>
<td>E12-6</td>
<td></td>
</tr>
<tr>
<td>Lecomte, P.</td>
<td>M2-10</td>
<td></td>
</tr>
<tr>
<td>Leggat, V.</td>
<td>C10-9</td>
<td></td>
</tr>
<tr>
<td>Legge, K.A.</td>
<td>D5-1</td>
<td></td>
</tr>
<tr>
<td>Legros, C.</td>
<td>G6-8</td>
<td></td>
</tr>
<tr>
<td>Lehr, A.</td>
<td>K6-4</td>
<td></td>
</tr>
<tr>
<td>Le Jossé, C.</td>
<td>G5-5</td>
<td></td>
</tr>
<tr>
<td>Le Lounan, R.</td>
<td>C3-5</td>
<td></td>
</tr>
<tr>
<td>Leroy, G.</td>
<td>G3-3</td>
<td></td>
</tr>
<tr>
<td>Lescure, G.</td>
<td>E1-1, E6-6</td>
<td></td>
</tr>
<tr>
<td>Lézenev, N.B.</td>
<td>G2-2, G2-3, G2-4</td>
<td></td>
</tr>
<tr>
<td>Li, M.</td>
<td>D1-2</td>
<td></td>
</tr>
<tr>
<td>Li, P.</td>
<td>C6-8</td>
<td></td>
</tr>
<tr>
<td>Lienard, P.A.</td>
<td>K1-1</td>
<td></td>
</tr>
<tr>
<td>Lim, M.K.</td>
<td>G5-7</td>
<td></td>
</tr>
<tr>
<td>Lin, W.W.</td>
<td>I1-4</td>
<td></td>
</tr>
<tr>
<td>Lindblom, B.</td>
<td>A6-4</td>
<td></td>
</tr>
<tr>
<td>Lindgren, F.</td>
<td>B11-6</td>
<td></td>
</tr>
<tr>
<td>Lindsay, D.S.</td>
<td>M1-9</td>
<td></td>
</tr>
<tr>
<td>Liu, X.</td>
<td>I3-1</td>
<td></td>
</tr>
<tr>
<td>Linanrees, J.</td>
<td>E10-6, E10-7, E11-6</td>
<td></td>
</tr>
<tr>
<td>Llopis, A.</td>
<td>E10-6, E10-7, E11-6</td>
<td></td>
</tr>
<tr>
<td>Long, C.R.</td>
<td>B7-5</td>
<td></td>
</tr>
<tr>
<td>Loo, D.</td>
<td>B6-2</td>
<td></td>
</tr>
<tr>
<td>Louwigné, B.</td>
<td>C4-5, E6-4</td>
<td></td>
</tr>
<tr>
<td>Lundström, R.</td>
<td>F2-2</td>
<td></td>
</tr>
<tr>
<td>Lunarto, E.</td>
<td>D6-7, E7-9, E12-6</td>
<td></td>
</tr>
<tr>
<td>Lyamshev, L.M.</td>
<td>G7-2</td>
<td></td>
</tr>
<tr>
<td>Lyon, R.H.</td>
<td>D3-6</td>
<td></td>
</tr>
<tr>
<td>Man, D.R.</td>
<td>K6-1, M3-1</td>
<td></td>
</tr>
<tr>
<td>Manville, P.</td>
<td>M6-6</td>
<td></td>
</tr>
<tr>
<td>Macecek, M.</td>
<td>H1-6</td>
<td></td>
</tr>
<tr>
<td>Mader, A.J.</td>
<td>J1-1</td>
<td></td>
</tr>
<tr>
<td>Maekawa, S.</td>
<td>E3-6</td>
<td></td>
</tr>
<tr>
<td>Mage, J.C.</td>
<td>D3-8, D6-1</td>
<td></td>
</tr>
<tr>
<td>Okujima, M.</td>
<td>D1-3, F1-2, F1-3, H1-3</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------------------</td>
<td></td>
</tr>
<tr>
<td>Odnett, A.M.</td>
<td>El2-3, El2-8</td>
<td></td>
</tr>
<tr>
<td>Oiza, J.</td>
<td>K3-1</td>
<td></td>
</tr>
<tr>
<td>Ojst, M.</td>
<td>B1-6</td>
<td></td>
</tr>
<tr>
<td>Örman, N.</td>
<td>C10-6</td>
<td></td>
</tr>
<tr>
<td>Oetiguy, G.L.</td>
<td>B10-7</td>
<td></td>
</tr>
<tr>
<td>Ostrovsky, L.A.</td>
<td>I3-10</td>
<td></td>
</tr>
<tr>
<td>Owana, J.</td>
<td>E3-2</td>
<td></td>
</tr>
<tr>
<td>Osselfeld, G.P.</td>
<td>C10-8, J1-8</td>
<td></td>
</tr>
<tr>
<td>Otsuka, T.</td>
<td>L2-3</td>
<td></td>
</tr>
<tr>
<td>Ottley, J.-P.</td>
<td>G7-1</td>
<td></td>
</tr>
<tr>
<td>Ovichiuk, V.</td>
<td>C2-2</td>
<td></td>
</tr>
<tr>
<td>Owana, S.</td>
<td>M5-7</td>
<td></td>
</tr>
<tr>
<td>Ox, N.</td>
<td>D6-4</td>
<td></td>
</tr>
<tr>
<td>Ozawa, H.</td>
<td>C2-1</td>
<td></td>
</tr>
</tbody>
</table>

| Painter, C.A. | J2-6 |
| Petisson, A. | E1-1-4 |
| Pandey, P.C. | A1-3 |
| Panuszka, R. | D3-10 |
| Paoloni, A. | A6-1-3, C1-2 |
| Papalardo, M. | L1-6 |
| Paré, L. | B5-6 |
| Pariah, J.R., J.R. | J2-1 |
| Park, S.H. | B8-3 |
| Parlow, U. | I4-7 |
| Parrot, J.-M. | D3-5 |
| Pauzé, S. | M2-10 |
| Pavilé, G. | D6-6 |
| Paves, M. | I3-6 |
| Pech, R. | C8-2, M2-3 |
| Pedersen, M.A. | H3-6 |
| Pegna, A. | B11-9 |
| Penner, A.R. | H2-3 |
| Perkel, J.S. | A3-5, A3-6, A3-10 |
| Perlin, R. | K2-8 |
| Peterson, K.W. | K2-8 |
| Petersson, R.A.T. | D6-3 |
| Petitjean, A. | C3-4, J3-9 |
| Petrovský, I. | G7-6 |
| Petrosky, J.S. | F2-1 |
| Petersson, O.Kr., Ø. | C5-4 |
| Pettorino, M. | A3-2 |
| Peutz, V.M.A. | E4-5 |
| Petermacher, J. | B10-8, L2-10 |
| Phanef, R. | B5-3, B5-8 |
| Phillips, C.A. | F2-1 |
| Phillips, I.B. | K5-5 |
| Plaud, J.-B. | M2-7 |
| Plaxa, R. | G3-5 |
| Pfeifer, A.D. | J2-1 |
| Pfeifer, J.E. | F2-4, F2-5, F2-6 |
| Pimomow, L. | M1-2 |
| Piraux, G. | C5-8 |
| Pirsa, P.M. | F2-7 |
| Planchard, J. | E7-9 |
| Plitnik, G. | M3-5 |
| Ploom, R. | B4-1, B11-4, E3-7 |
| Plump, J.M. | E4-1 |
| Polack, J.-D. | E5-3, E6-4 |
| Pons, J. | G10-7 |
| Pope, J. | M3-3 |
| Popplewell, N. | D5-2 |
| Piasetzki, C. | H1-6 |
| Potteat, T.L. | L1-2 |
| Potter, D. | H2-7 |
| Pouliquen, J. | G5-2, G5-4 |
| Poulsen, T. | B9-4 |
| Powell, R.L. | I2-6 |
| Prang, L.C.S. | C10-1 |

| Prasad, M.G. | M3-6 |
| Price, I.R. | F2-7 |
| Fritz, T. | D5-9 |
| Prosperetti, A. | I7-6 |
| Purshouse, M. | C2-4 |
| Fyukk, I. | F2-3 |

| Quentin, C.J. | C3-2, L1-4 |

| Reastgeber, J. | B2-4 |
| Rabbitt, R.D. | B7-4 |
| Rajan, S. | D5-4, D5-5 |
| Kamakrishna, B.S. | K1-7 |
| Ranachowski, J. | G6-6 |
| Rasmussen, C. | M2-6 |
| Rasmussen, F. | M2-3 |
| Recuero, M. | G6-9, M4-8 |
| Reddy, K.G. | K1-7 |
| Richman, J. | C3-6 |
| Reif, G.N. | H1-2 |
| Reif, Z.F. | F2-5 |
| Reijmund, F. | G6-6 |
| Renew, W.D. | C9-1 |
| Rheault, M. | C7-9 |
| Ribner, H.H. | J3-2 |
| Rice, K. | A6-3 |
| Rich, N.C. | B7-7 |
| Richards, T.L. | I5-3 |
| Richardson, B.K. | K6-6 |
| Richter, D. | B9-2 |
| Riera Franco de S., E. | G6-2 |
| Ripoche, J. | G1-3, I1-3, M1-4 |
| Roberts, G.W. | K3-1 |
| Robertson, A.W. | H3-4 |
| Robinson, D.W. | B11-1 |
| Robles, L. | B7-7 |
| Rodriguez, G. | G4-2 |
| Roebroek, J.G.H. | D3-5 |
| Rogers, P.H. | H6-7 |
| Rokas, D. | B10-3 |
| Rokhin, J.I. | G3-4 |
| Roland, J. | E1-4 |
| Ross, F. | C3-5, C3-7 |
| Rowe, J.H. | G3-5 |
| Rosenhouse, G. | G8-3 |
| Rossetto, B. | D2-1 |
| Rossi, M. | A2-7 |
| Rossing, T.D. | B2-1, K2-8 |
| Roug, L. | A6-4 |
| Rouquié, G. | A3-5 |
| Roure, A. | C5-5 |
| Roux, J.E. | Plenary 4 |
| Rozeba, E.C. | E11-1 |
| Roszynski, A.J. | A2-4 |
| Ruggero, M.A. | B7-7 |
| Ruske, G. | A1-5 |

<p>| Sahr, R. | B4-6 |
| Saikawa, K. | A1-3 |
| Saito, T. | A4-5 |
| Sakakibara, H. | K3-8 |
| Sakurai, Y. | B3-1, C1-4 |
| Salamon, K. | L5-3 |
| Salazar, E.R. | B10-2 |
| Salza, F.L. | A6-2 |
| Sanada, H. | D3-1 |
| Sánchez, L. | L2-10 |</p>
<table>
<thead>
<tr>
<th>Name</th>
<th>Seat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sancho, J.</td>
<td>E10-6, E10-7, E11-6</td>
</tr>
<tr>
<td>Sandri, S.</td>
<td>A6-2</td>
</tr>
<tr>
<td>San Emanterio, J.L.</td>
<td>G4-2</td>
</tr>
<tr>
<td>Sana, K.</td>
<td>G5-8, G6-1</td>
</tr>
<tr>
<td>Santerre, L.</td>
<td>A2-2</td>
</tr>
<tr>
<td>Santiago, J.S.</td>
<td>C10-2</td>
</tr>
<tr>
<td>Sasaki, M.</td>
<td>C1-2</td>
</tr>
<tr>
<td>Sathoff, M.J.</td>
<td>K2-8</td>
</tr>
<tr>
<td>Satory, R.J.</td>
<td>E5-5</td>
</tr>
<tr>
<td>Savage, W.S.</td>
<td>K2-4</td>
</tr>
<tr>
<td>Sawan, M.</td>
<td>B3-3</td>
</tr>
<tr>
<td>Schaefer, P.</td>
<td>C1-3</td>
</tr>
<tr>
<td>Scherer, F.</td>
<td>L2-9</td>
</tr>
<tr>
<td>Schick, A.</td>
<td>C1-1</td>
</tr>
<tr>
<td>Schmidt, H.</td>
<td>M2-5</td>
</tr>
<tr>
<td>Schmutz, J.</td>
<td>C2-7</td>
</tr>
<tr>
<td>Schneider, T.</td>
<td>C7-4</td>
</tr>
<tr>
<td>Schoenberg, M.</td>
<td>I5-5</td>
</tr>
<tr>
<td>Schreiber, L.</td>
<td>C10-4, C10-5</td>
</tr>
<tr>
<td>Schoot, J.</td>
<td>A4-4, B1-6, B10-5</td>
</tr>
<tr>
<td>Schultz, O.</td>
<td>M2-1</td>
</tr>
<tr>
<td>Schultz, T.J.</td>
<td>E4-4</td>
</tr>
<tr>
<td>Schumacher, R.T.</td>
<td>K3-3, K3-6</td>
</tr>
<tr>
<td>Scory, N.</td>
<td>P2-8</td>
</tr>
<tr>
<td>Seidel, K.</td>
<td>K11-5</td>
</tr>
<tr>
<td>Sermerecigil, E.</td>
<td>D5-2</td>
</tr>
<tr>
<td>Sen, F.N.</td>
<td>I5-5</td>
</tr>
<tr>
<td>Sen, K.</td>
<td>I5-2</td>
</tr>
<tr>
<td>Serren, C.</td>
<td>D3-9</td>
</tr>
<tr>
<td>Serignat, J.F.</td>
<td>A4-7</td>
</tr>
<tr>
<td>Sessler, G.M.</td>
<td>G7-5</td>
</tr>
<tr>
<td>Sette, D.</td>
<td>D5-7</td>
</tr>
<tr>
<td>Seyer, K.</td>
<td>L2-3, L2-6</td>
</tr>
<tr>
<td>Seybert, A.F.</td>
<td>M2-2</td>
</tr>
<tr>
<td>She, J.</td>
<td>C5-6</td>
</tr>
<tr>
<td>Shadle, C.H.</td>
<td>A3-4</td>
</tr>
<tr>
<td>Shakharpanov, M.</td>
<td>G2-2</td>
</tr>
<tr>
<td>Shanov, A.A.</td>
<td>G2-3</td>
</tr>
<tr>
<td>Shang, W.C.</td>
<td>I3-9</td>
</tr>
<tr>
<td>Shanker, K.</td>
<td>G1-5</td>
</tr>
<tr>
<td>Shao, L.</td>
<td>E10-4</td>
</tr>
<tr>
<td>Shaw, K.A.G.</td>
<td>B6-6</td>
</tr>
<tr>
<td>Shaw, K.A.</td>
<td>G9-5</td>
</tr>
<tr>
<td>Shen, J.H.</td>
<td>I3-6</td>
</tr>
<tr>
<td>Shen, Y.</td>
<td>B11-2</td>
</tr>
<tr>
<td>Shihayama, H.</td>
<td>L2-5</td>
</tr>
<tr>
<td>Shikada, N.</td>
<td>J1-5</td>
</tr>
<tr>
<td>Shima, N.</td>
<td>H1-1</td>
</tr>
<tr>
<td>Shimada, S.</td>
<td>D3-2</td>
</tr>
<tr>
<td>Shimomura, Y.</td>
<td>D3-6</td>
</tr>
<tr>
<td>Shipton, M.S.</td>
<td>B9-3</td>
</tr>
<tr>
<td>Shirai, K.</td>
<td>H1-4</td>
</tr>
<tr>
<td>Shirasuna, S.</td>
<td>B6-3</td>
</tr>
<tr>
<td>Shuai, Z.</td>
<td>B11-2</td>
</tr>
<tr>
<td>Shui, Y.</td>
<td>G1-2</td>
</tr>
<tr>
<td>Stibilla, C.</td>
<td>D5-7</td>
</tr>
<tr>
<td>Sdiki, M.</td>
<td>C8-6, J1-6</td>
</tr>
<tr>
<td>Slabek, G.W.</td>
<td>K5-2, E11-1</td>
</tr>
<tr>
<td>Silverberg, L.M.</td>
<td>D4-4</td>
</tr>
<tr>
<td>Simon, H.J.</td>
<td>B4-2</td>
</tr>
<tr>
<td>Sinha, D.K.</td>
<td>C1-4</td>
</tr>
<tr>
<td>Slawinska, E.B.</td>
<td>A2-5</td>
</tr>
<tr>
<td>Slawinski, A.</td>
<td>D1-7</td>
</tr>
<tr>
<td>Smoak, P.</td>
<td>C8-4, D3-3</td>
</tr>
<tr>
<td>Smith, E.D.</td>
<td>G5-10</td>
</tr>
<tr>
<td>Smith, T.J.B.</td>
<td>D1-5</td>
</tr>
<tr>
<td>Smorencburg, C.F.</td>
<td>B9-6</td>
</tr>
<tr>
<td>Smotryszak, J.</td>
<td>B1-7</td>
</tr>
<tr>
<td>Soetanto, K.</td>
<td>F1-2, F1-3</td>
</tr>
<tr>
<td>Sofer, J.M.</td>
<td>G6-3, G6-5</td>
</tr>
<tr>
<td>Solana-Quiros, P.</td>
<td>C4-7</td>
</tr>
<tr>
<td>Sommerfeldt, S.D.</td>
<td>K4-3</td>
</tr>
<tr>
<td>Sondhi, H.M.</td>
<td>A4-4</td>
</tr>
<tr>
<td>Sone, T.</td>
<td>B2-10, B11-10</td>
</tr>
<tr>
<td>Sørstrad, S.</td>
<td>C5-6</td>
</tr>
<tr>
<td>Soufransky, J.</td>
<td>K4-6</td>
</tr>
<tr>
<td>Spencer, N.A.</td>
<td>K2-2</td>
</tr>
<tr>
<td>Stewick, H.</td>
<td>J3-7</td>
</tr>
<tr>
<td>Stinson, N.K.</td>
<td>B6-4, B6-6</td>
</tr>
<tr>
<td>Stroynow, J.</td>
<td>D6-5</td>
</tr>
<tr>
<td>Strassberger, W.</td>
<td>J1-3</td>
</tr>
<tr>
<td>Strassewicz, W.</td>
<td>E12-9</td>
</tr>
<tr>
<td>Strze, S.</td>
<td>E4-7</td>
</tr>
<tr>
<td>Strong, W.J.</td>
<td>K4-3</td>
</tr>
<tr>
<td>Sugawara, S.</td>
<td>D1-1, M4-7</td>
</tr>
<tr>
<td>Sugiyama, T.</td>
<td>E11-5</td>
</tr>
<tr>
<td>Sui, X.</td>
<td>K1-6</td>
</tr>
<tr>
<td>Sutin, A.M.</td>
<td>I3-10</td>
</tr>
<tr>
<td>Suzuki, H.</td>
<td>L3-1</td>
</tr>
<tr>
<td>Suzuki, S.</td>
<td>J3-8</td>
</tr>
<tr>
<td>Suzuki, T.</td>
<td>J1-5</td>
</tr>
<tr>
<td>Suzuki, Y.</td>
<td>B2-10</td>
</tr>
<tr>
<td>Swean, J.</td>
<td>E9-1</td>
</tr>
<tr>
<td>Swallow, J.C.</td>
<td>D1-4</td>
</tr>
<tr>
<td>Swanson, D.C.</td>
<td>D4-3</td>
</tr>
</tbody>
</table>

- T -

<table>
<thead>
<tr>
<th>Name</th>
<th>Seat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tachibana, K.</td>
<td>C8-7, C8-8, J1-4</td>
</tr>
<tr>
<td>Talbo, L.M.</td>
<td>C10-10</td>
</tr>
<tr>
<td>Teelal, D.</td>
<td>G5-5</td>
</tr>
<tr>
<td>Takagi, K.</td>
<td>G1-1, G2-1</td>
</tr>
<tr>
<td>Takahashi, B.</td>
<td>E7-6</td>
</tr>
<tr>
<td>Takahashi, M.</td>
<td>C5-1, C5-3</td>
</tr>
<tr>
<td>Takano, Y.</td>
<td>D3-2</td>
</tr>
<tr>
<td>Takada, K.</td>
<td>C4-2</td>
</tr>
<tr>
<td>Tanaka, M.</td>
<td>I5-1</td>
</tr>
<tr>
<td>Tanahe, K.</td>
<td>C3-2</td>
</tr>
<tr>
<td>Tanahe, Shogo</td>
<td>D3-1</td>
</tr>
<tr>
<td>Tanahe, Souei</td>
<td>E7-5</td>
</tr>
<tr>
<td>Tanaka, H.</td>
<td>B11-10</td>
</tr>
<tr>
<td>Tanahe, M.</td>
<td>F1-3</td>
</tr>
<tr>
<td>Taudara, V.</td>
<td>C2-8, C2-10</td>
</tr>
<tr>
<td>Tango, G.J.</td>
<td>I1-7</td>
</tr>
<tr>
<td>Tao, J.</td>
<td>H3-7</td>
</tr>
<tr>
<td>Taylor, W.</td>
<td>F2-4, F2-5, F2-6</td>
</tr>
<tr>
<td>TenCate, J.A.</td>
<td>C6-2</td>
</tr>
<tr>
<td>Ten Kate, J.H.</td>
<td>B8-5</td>
</tr>
<tr>
<td>Ten Kao, B.H.</td>
<td>B9-6</td>
</tr>
<tr>
<td>Terasaki, R.</td>
<td>K2-3</td>
</tr>
<tr>
<td>Teysou, R.</td>
<td>B6-2</td>
</tr>
<tr>
<td>Thierry, L.</td>
<td>B5-4</td>
</tr>
<tr>
<td>Thomas, R.</td>
<td>A5-1</td>
</tr>
<tr>
<td>Thompson, G.</td>
<td>I2-6, I3-2, J3-3</td>
</tr>
<tr>
<td>Thompson, W.F.</td>
<td>K5-3</td>
</tr>
<tr>
<td>Thorleifson, J.M.</td>
<td>W2-3</td>
</tr>
<tr>
<td>Thorpe, L.A.</td>
<td>K5-2</td>
</tr>
<tr>
<td>Tien, J.</td>
<td>G9-6</td>
</tr>
<tr>
<td>Tichy, J.</td>
<td>D4-3</td>
</tr>
<tr>
<td>Tien, H.F.</td>
<td>L1-2</td>
</tr>
<tr>
<td>Tietz, J.R.</td>
<td>I2-4</td>
</tr>
<tr>
<td>Tipton, S.</td>
<td>I2-4</td>
</tr>
<tr>
<td>To, C.W.S.</td>
<td>G6-1</td>
</tr>
<tr>
<td>Tomita, S.</td>
<td>C3-2</td>
</tr>
<tr>
<td>Torhara, H.</td>
<td>E9-4</td>
</tr>
<tr>
<td>Torscano, M.</td>
<td>A3-2</td>
</tr>
<tr>
<td>Tousatier, M.</td>
<td>D5-8</td>
</tr>
<tr>
<td>Toyota, Y.</td>
<td>E6-1</td>
</tr>
<tr>
<td>Treher, E.L.</td>
<td>K5-2, K5-4</td>
</tr>
<tr>
<td>Trethewey, M.W.</td>
<td>N3-7</td>
</tr>
<tr>
<td>Tro, J.</td>
<td>E6-7</td>
</tr>
<tr>
<td>Truchon-Gagnon, C.</td>
<td>C1-5</td>
</tr>
<tr>
<td>Tse, S.-M.</td>
<td>A2-3</td>
</tr>
<tr>
<td>Tsui, C.Y.</td>
<td>H1-2</td>
</tr>
<tr>
<td>Tubach, J.P.</td>
<td>A6-4</td>
</tr>
<tr>
<td>Tubits, A.</td>
<td>B7-5, K2-7</td>
</tr>
<tr>
<td>Name</td>
<td>Code</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
</tr>
<tr>
<td>Tuffelli, D.</td>
<td>BB-4</td>
</tr>
<tr>
<td>Tweed, J.J.</td>
<td>B7-2</td>
</tr>
<tr>
<td>Übler, H.</td>
<td>D6-5, M1-3</td>
</tr>
<tr>
<td>Ueba, S.</td>
<td>D6-8</td>
</tr>
<tr>
<td>Uehata, T.</td>
<td>E6-2</td>
</tr>
<tr>
<td>Ugli, Y.</td>
<td>J3-8</td>
</tr>
<tr>
<td>Van Bokum, M.F.</td>
<td>B8-5</td>
</tr>
<tr>
<td>Van Chi, N.</td>
<td>E7-9</td>
</tr>
<tr>
<td>Van den Berg, F.H.A.</td>
<td>E6-5</td>
</tr>
<tr>
<td>Van Dijkhuysen, J.N.</td>
<td>B4-1</td>
</tr>
<tr>
<td>Van Eercke, D.</td>
<td>E11-7</td>
</tr>
<tr>
<td>van Riel, P.</td>
<td>G4-4</td>
</tr>
<tr>
<td>Vass, T.</td>
<td>E2-6</td>
</tr>
<tr>
<td>Ventura, J.C.</td>
<td>B11-9</td>
</tr>
<tr>
<td>Verbrugghe, E.</td>
<td>G5-8, G6-1</td>
</tr>
<tr>
<td>Vian, J.P.</td>
<td>B4-10</td>
</tr>
<tr>
<td>Villacce, G.</td>
<td>A3-4</td>
</tr>
<tr>
<td>Villouvier, V.</td>
<td>C6-4</td>
</tr>
<tr>
<td>Voroncov, N.N.</td>
<td>C6-5</td>
</tr>
<tr>
<td>Vos, J.</td>
<td>K5-8</td>
</tr>
<tr>
<td>Vray, D.</td>
<td>G1-6</td>
</tr>
<tr>
<td>Wagstaff, F.R.</td>
<td>C8-2, C8-5, M2-3</td>
</tr>
<tr>
<td>Wakita, H.</td>
<td>A4-8</td>
</tr>
<tr>
<td>Walker, G.P.</td>
<td>K3-7</td>
</tr>
<tr>
<td>Walker, R.</td>
<td>D1-5</td>
</tr>
<tr>
<td>Wallis, A.D.</td>
<td>M4-1, M4-2</td>
</tr>
<tr>
<td>Walsh, J.P.</td>
<td>E5-7</td>
</tr>
<tr>
<td>Wang, B.</td>
<td>H3-7</td>
</tr>
<tr>
<td>Wang, N.</td>
<td>C6-8</td>
</tr>
<tr>
<td>Wang, J.</td>
<td>E10-4</td>
</tr>
<tr>
<td>Wang, S.</td>
<td>K1-3</td>
</tr>
<tr>
<td>Wang, Y.Y.</td>
<td>H3-9</td>
</tr>
<tr>
<td>Wang, Z.</td>
<td>E5-6</td>
</tr>
<tr>
<td>Wang, Zhiguo</td>
<td>C6-3</td>
</tr>
<tr>
<td>Ward, W.D.</td>
<td>B3-2</td>
</tr>
<tr>
<td>Warschaw, S.I.</td>
<td>L3-6</td>
</tr>
<tr>
<td>Waszkel, O.</td>
<td>E5-4</td>
</tr>
<tr>
<td>Wasserman, D.E.</td>
<td>F2-1</td>
</tr>
<tr>
<td>Watanabe, T.</td>
<td>D1-3</td>
</tr>
<tr>
<td>Waterhouse, R.V.</td>
<td>G4-6</td>
</tr>
<tr>
<td>Wei, R.</td>
<td>H3-7</td>
</tr>
<tr>
<td>Weigel, W.</td>
<td>A1-5</td>
</tr>
<tr>
<td>Weinreich, G.</td>
<td>K3-7</td>
</tr>
<tr>
<td>Weiss, W.</td>
<td>A5-1</td>
</tr>
<tr>
<td>Werby, M.F.</td>
<td>D6-5, I1-6, I1-7</td>
</tr>
<tr>
<td>Werner, E.</td>
<td>B4-5, B4-6</td>
</tr>
<tr>
<td>West, C.D.</td>
<td>B8-1</td>
</tr>
<tr>
<td>West, J.</td>
<td>G7-5</td>
</tr>
<tr>
<td>Weertleda, A.</td>
<td>E11-4</td>
</tr>
<tr>
<td>Witt, F.</td>
<td>C8-9, C8-10</td>
</tr>
<tr>
<td>Wheeler, F.D.</td>
<td>E10-4</td>
</tr>
<tr>
<td>Whitaker, R.W.</td>
<td>J2-6</td>
</tr>
<tr>
<td>Wilken, W.</td>
<td>J1-10</td>
</tr>
<tr>
<td>Williams, K.L.</td>
<td>I1-2</td>
</tr>
<tr>
<td>Woodcock, R.</td>
<td>E7-1</td>
</tr>
<tr>
<td>Wooten, R.F.</td>
<td>I1-5</td>
</tr>
<tr>
<td>Wright, K.D.</td>
<td>A4-3</td>
</tr>
<tr>
<td>Wu, H.Q.</td>
<td>C8-5</td>
</tr>
<tr>
<td>Xi, D.</td>
<td>E8-7</td>
</tr>
<tr>
<td>Xiang, F.N.</td>
<td>E3-6</td>
</tr>
</tbody>
</table>

- Y -
<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yadav, R.R.</td>
<td>G2-8</td>
</tr>
<tr>
<td>Yamamoto, S.</td>
<td>M4-4</td>
</tr>
<tr>
<td>Yamashita, Y.</td>
<td>B1-4</td>
</tr>
<tr>
<td>Yamada, M.</td>
<td>A1-6</td>
</tr>
<tr>
<td>Yano, H.</td>
<td>C8-1, C8-8</td>
</tr>
<tr>
<td>Yano, K.</td>
<td>M4-4</td>
</tr>
<tr>
<td>Yao, D.</td>
<td>D6-2</td>
</tr>
<tr>
<td>Yatsumi, T.</td>
<td>M4-4</td>
</tr>
<tr>
<td>Yeates, E.M.</td>
<td>H1-5</td>
</tr>
<tr>
<td>Ying, C.F.</td>
<td>I3-6</td>
</tr>
<tr>
<td>Yoneyama, T.</td>
<td>E10-9</td>
</tr>
<tr>
<td>Yoo, R.K.</td>
<td>B8-3</td>
</tr>
<tr>
<td>Yoshidai, K.</td>
<td>J1-4</td>
</tr>
<tr>
<td>Yoshikawa, S.</td>
<td>M4-4</td>
</tr>
<tr>
<td>Yost, W.T.</td>
<td>G3-1</td>
</tr>
<tr>
<td>Young, L.H.</td>
<td>J2-3</td>
</tr>
<tr>
<td>Youn, R.W.</td>
<td>J2-7</td>
</tr>
<tr>
<td>Yucase, A.</td>
<td>G5-5</td>
</tr>
<tr>
<td>Yund, E.W.</td>
<td>B4-2</td>
</tr>
</tbody>
</table>

- Z -
<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zagrai, N.</td>
<td>D5-7</td>
</tr>
<tr>
<td>Zhang, D.</td>
<td>G1-2</td>
</tr>
<tr>
<td>Zhang, J.</td>
<td>B11-2</td>
</tr>
<tr>
<td>Zhang, M.I.</td>
<td>E5-6</td>
</tr>
<tr>
<td>Zhang, R.</td>
<td>H3-5</td>
</tr>
<tr>
<td>Zhang, S.Y.</td>
<td>I3-6</td>
</tr>
<tr>
<td>Zhao, S.L.</td>
<td>C6-5</td>
</tr>
<tr>
<td>Zheng, X.</td>
<td>H3-7</td>
</tr>
<tr>
<td>Zhou, G.</td>
<td>E2-3</td>
</tr>
<tr>
<td>Zhu, C.</td>
<td>I3-1</td>
</tr>
<tr>
<td>Zhu, Z.</td>
<td>C6-3</td>
</tr>
<tr>
<td>Zindel, M.</td>
<td>J3-5</td>
</tr>
<tr>
<td>Zinskinov, V.V.</td>
<td>G7-2</td>
</tr>
<tr>
<td>Zucker, A.J.</td>
<td>J2-4</td>
</tr>
<tr>
<td>Zinkier, K.</td>
<td>L2-9</td>
</tr>
<tr>
<td>Zwolicki, J.J.</td>
<td>B7-6</td>
</tr>
</tbody>
</table>

- X -
<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xi, D.</td>
<td>E8-7</td>
</tr>
<tr>
<td>Xiang, F.N.</td>
<td>E3-6</td>
</tr>
</tbody>
</table>