SPEAKER IDENTIFICATION USING NONCONTEMPORARY SPEECH SAMPLES

Harry Hollien and Reva Schwartz

Institute for Advanced Study of the Communication Processes
University of Florida, Gainesville, FL 32611

Abstract: Two noncontemporary elements are important to the aural-perceptual speaker identification process; either 1) listeners are asked to make identifications at various times after hearing the speaker or 2) samples of the speaker's voice are obtained at different points in time. It has been found that memory for voice decays rapidly in the first case; hence it is assumed that the same is true when noncontemporary samples are used. To test this second relationship, samples with latencies of 4, 8 and 32 weeks, and, then later, 6 and 20 years were studied. Mean correct identification was initially reduced by almost 20% and this level of correct identification was sustained for up to 6 years; it dropped significantly only for the 20-year latencies.

INTRODUCTION

The two noncontemporary factors cited in the abstract (delays in hearing the talker's speech and speech samples taken at different times) are among the many variables that affect the efficiency and validity of aural-perceptual speaker identification. Research on the first of these issues can be used to demonstrate that auditors are not very adept at recognizing the voices of unfamiliar speakers and that even these levels of correct identification decay sharply over time. While current opinion supports the notion that the use of noncontemporary speech samples will result in similar degradation of correct identification levels, other logic can be used to argue the contrary. Only two investigators (Rothman, 1977; Schwartz, 1995) have researched this issue and since their results are sharply in conflict with each other, it would appear useful to further examine the relationship.

METHOD

This research to follow was carried out in two parts. First an experiment that duplicated and extended those reported by the cited authors was completed. Then, due to the obtained results, a second was carried out; one which extended the research over a substantial period of time. The two studies have been combined in this report. Both involved the ability of auditors to identify male talkers from paired samples where each of the tokens were made at different times. The procedure used was to present the pairs in two sets of subexperiments; in the first pair, the latencies were four and eight weeks and then four and 32 weeks. The second set involved even longer periods; here the pairs for the third study were recorded six years apart and then those for the final one, 20 years apart.

Either 10 or 11 young, healthy male talkers were used in each of the four procedures. In the first two they read complex sentences 7-9 sec. in length; in the second two (6 and 20 years) 7-9 sec. sentences were drawn from a standard paragraph. In all instances the talker was paired with himself. Finally, four groups of listeners were employed; with 31 to 41 young healthy university students in each cohort. All had to demonstrate good hearing and that they could correctly identify like and different voices at an 85% level or better. These "test" samples were embedded randomly among those which make up the overall experimental tape.

RESULTS

The results of both studies are summarized in Table 1. Note first the selection criteria; all subjects comfortably exceeded the criterion levels (85%). Second, the data in the "same" column doubles as the baseline or contemporary reference. That is, auditors can be expected to correctly indicate whether or not both talkers are the same person (ABX procedure) if the voices for that pair of samples are produced by a single individual. In short, these data are for contemporary samples. Third, if the results of the short
term latencies (i.e., 4-32 weeks) are taken as a whole, it may be seen that a general drop of (about) an additional 20% can be expected when the noncontemporary samples are compared to the contemporary.

Table 1. Summary table of listeners responses to noncontemporary speech samples (4 weeks-20 years) uttered by male talkers. All scores, except number of subjects, are in percent. The NC1 and NC2 data are for the first pair of studies whereas the NC3, NC4 are for the second pair.

<table>
<thead>
<tr>
<th>Study</th>
<th>Listeners</th>
<th>Selection Criteria</th>
<th>Correct Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Talkers</td>
<td>same</td>
<td>different</td>
</tr>
<tr>
<td></td>
<td></td>
<td>91.2</td>
<td>96.3</td>
</tr>
<tr>
<td>NC1</td>
<td>36</td>
<td>10</td>
<td>94.0</td>
</tr>
<tr>
<td>NC2</td>
<td>31</td>
<td>11</td>
<td>96.6</td>
</tr>
<tr>
<td>NC3</td>
<td>41</td>
<td>10</td>
<td>97.9</td>
</tr>
<tr>
<td>NC4</td>
<td>41</td>
<td>11</td>
<td>95.1</td>
</tr>
</tbody>
</table>

The most striking finding from the first pair of studies is that they strongly support the data reported by Schwartz (1995) and are substantially better than those reported earlier by Rothman (1977). Thus, the logic that a speaker should sound like himself, even if the recordings of his voice are made at two different points in time, was supported and the illusion that degradation will occur even if memory is not a factor was not. Finally, the second set if experiments (latencies of 6 and 20 years) are quite consistent with the short term findings. Indeed, the talkers were almost easier to identify after the 6-yr. delay. It was not until 20 years separated the two samples that the correct identification scores were reduced to rather low levels -- a condition that probably resulted from changes in speakers' motor speech capabilities over that long period of time.

CONCLUSION

A marked reduction, then a rapid further decay, in the ability of listeners to identify individuals from noncontemporary speech samples has been suggested; however, the data obtained in this research did not support that postulation. Hence, it can be concluded that noncontemporary speech samples can be expected to show only minimal degradation (for aural-perceptual speaker identification purposes anyway), and do so for periods of up to six years, and perhaps even longer. Thus, it would appear that the use of noncontemporary utterances will have but a minimal effect on the accuracy of this type of speaker identification.

REFERENCES