An evaluation method of coded signal by means of Frequency Domain Binaural Model

Haruki Morinaga*, Satoshi Orita*, Takashi Nakanishi*, Yoshifumi Chisaki**, Tsuyoshi Usagawa**

*Graduate School of Science and Technology, Kumamoto University
{mokkun,satoshi,takashin}@cogni.cs.kumamoto-u.ac.jp
**Department of Computer Science, Faculty of Engineering, Kumamoto University
{chisaki,tuie}@cs.kumamoto-u.ac.jp

Abstract

When performance of coding algorithm is evaluated, we often experience the difference between impression of audio demonstration and a physical index such as SNR. This difference might be due to the fact that SNR does not take account of masking effects and loudness. Then, evaluation method based on a psychoacoustic model used in MPEG1/Audio was proposed previously. However, this method can apply only for monaural signal of each channel of stereo signal independently. In case of taking account of binaural effects, perception of sound localization is an important factor for evaluating stereo signal such as enhanced signal, compressed speech, and coded signal. In this paper, an evaluation method based on a psychoacoustic model for stereo signal is proposed using frequency domain binaural model (FDBM). This method utilizes both an interaural level difference (ILD) and an interaural phase difference (IPD) of stereo signal. Details of this method as follows: At first, an original and a target stereo signals are analyzed using FDBM separately. FDBM provides direction-of-arrival (DOA) information of each frequency bin based on the ILD and IPD. And a masking threshold of each frequency bin is obtained in order to pick up audible components. Based on DOA and threshold, each stereo signal can be represented as a contour map against DOA and frequency axes. The proposed evaluation method compares the maps for the original and the target signals.

1. Introduction

Efficient signal coding method takes account of human auditory property known as masking effects. MPEG1/Audio[1] is one of the standardized signal compression methods. When performance of such signal coding or signal enhancement methods is evaluated, we often experience the difference between impression of audio demonstration and a physical index such as SNR. This difference might be due to the fact that SNR does not consider masking effects and loudness. Previously, an evaluation method based on a psychoacoustic model used in MPEG1/Audio was proposed [2]. Although this method is effective, it can apply only for monaural signal or each channel of stereo signal independently. In case of taking account of binaural effects, perception of sound localization is an important factor for evaluating stereo signal. In this paper, an evaluation method based on a psychoacoustic model for stereo signal is proposed using frequency domain binaural model (FDBM)[3]. This method utilizes both an interaural level difference (ILD) and an interaural phase difference (IPD) of stereo signal. The proposed method is examined by simulations and subjective tests.

2. Evaluation method

The block diagram of the proposed method is shown in Fig.1(a). Figure 1(b) shows the detail of auditory model which takes account of binaural effects using FDBM and monaural psychoacoustic model. Outputs of model for original and coded signals are obtained by this auditory model. The differences of outputs between original and coded signals are used for evaluation. Auditory model is composed of three blocks; DOA estimation using FDBM, calculation of masking threshold based on a psychoacoustic model used in MPEG1/Audio[1], and selection of DOA in each frequency components with taking masking effects into consideration.

2.1. DOA estimation using FDBM

FDBM is a signal enhancement method using DOA information which utilizes both an interaural level difference (ILD) and an interaural phase difference (IPD) of stereo signal[3]. The proposed evaluation method uses this DOA estimation process of FDBM for getting information of sound localization.

At first, left and right channel signals, \( l(t) \) and \( r(t) \), are transformed into spectra, \( X_L(k) \) and \( X_R(k) \), by FFT. Then cross-spectrum \( C_{lr}(k) \) is calculated from those spectra as follows:

\[
C_{lr}(k) = X_L(k)X_R^*(k),
\]

where \( * \) denotes a complex conjugate. When the power spectrum is expressed as \( C_{ll}(k) \), interaural level differ-
ence is defined as
\[ \xi_{tr}(k) = 20 \log \left| \frac{C_{tr}(k)}{C_{tr}(k)} \right|. \]  
(2)

Interaural phase difference is also defined as
\[ \phi_{tr}(k) = \tan^{-1} \left( \frac{\Im(C_{tr}(k))}{\Re(C_{tr}(k))} \right). \]  
(3)

Then, DOA information \( \hat{\theta}(k) \) of each frequency component is determined by comparing ILD with ILD-DOA map for the higher frequency range more than 1500Hz, IPD with IPD-DOA map for the lower frequency range less than 750Hz, and both IPD and ILD with IPD-DOA and ILD-DOA maps for the frequency range between 750 Hz and 1500 Hz. The HRTF database of KEMAR dummy-head microphone, which is provided by MIT Media Lab[4], is utilized as IPD-DOA and ILD-DOA maps in this implementation.

### 2.2. Calculation of masking threshold

Masking thresholds of original and coded signals are derived by means of a psychoacoustic model used in MPEG1/Audio[1], which is modified by authors[2]. The evaluation method based on a psychoacoustic model is performed both on linear distortion (LD) and noise including nonlinear distortion (NND); linear distortion is evaluated by the level difference between original and coded signals at which frequency the signal level is higher than the masking threshold, and noise including nonlinear distortion is evaluated by the spectral level of enhanced signal at which frequency the signal level is higher than the masking threshold.

**2.3. Decision of available DOA information**

Conditions of DOA estimation are shown in Fig.2. The audible signal components are considered as target signal. DOA information corresponding to human auditory spectra is decided by the following expression,

\[ \hat{\theta}(k) = \begin{cases} \theta(k) : X_L(k) \geq T_L(k) \& X_R(k) \geq T_R(k), \\ -90^\circ : X_L(k) \geq T_L(k) \& X_R(k) \leq T_R(k), \\ 90^\circ : X_L(k) \leq T_L(k) \& X_R(k) \geq T_R(k), \\ NA : X_L(k) \leq T_L(k) \& X_R(k) \leq T_R(k) \end{cases} \]  
(4)

where NA is not available, \( X_L(k) \) and \( X_R(k) \) are power spectral component of left and right signals, respectively. Masking thresholds of signals are shown as \( T_L(k) \) and \( T_R(k) \). When power spectral components of both channels are higher than masking threshold, DOA information calculated by FDBM is used as it is. If the component of either one of right or left channel is lower than masking threshold, then the DOA information is replaced according to expression condition as \( \hat{\theta}(k) = \pm 90^\circ \). When components of both channels are lower than masking threshold, then the DOA information is set to “Not-available.” Power spectrum of each frequency index \( k \) with DOA information \( P(\hat{\theta}, k) \) is obtained by these processes of auditory model.

### 2.4. Evaluation process

Calculated power spectrums with DOA information corresponding to an original and a coded signals are defined as \( P(\hat{\theta}, k) \) and \( P'(\hat{\theta}, k) \). An evaluation value \( eval(\theta, k) \) based on a psychoacoustic model at frequency index \( k \) and DOA = \( \theta \) is defined by the following expression,

\[ eval(\theta, k) = |P(\hat{\theta}, k) - P'(\hat{\theta}, k)|. \]  
(5)
3. Simulations

In this section, the results of evaluation using the proposed method are shown. A cappella music is selected as input stereo signal, and its compressed signals by various bit rate are used for evaluation. The sampling frequency of the signal is 44.1 kHz, and the quantization is 16 bit. Compression method is MPEG1/Audio Layer-3 (MP3), and the bit rate is set to 128 kbps, 96 kbps, 80 kbps, or 64 kbps of joint stereo mode.

3.1. Linear Distortion (LD) evaluation

The obtained spectrograms evaluated by the previous monaural evaluation method in each bit rate are shown in Fig. 3. Those signals are processed only left channel signal. The vertical and the horizontal axis shows frequency and time, respectively. The shading shows the power level of linear distortion, where brighter part corresponds to distorted part of signal.

According to the results, there are no apparent differences between 128 kbps and 96 kbps. However, the bit rate becomes lower than 80 kbps, linear distortion is shown in higher frequency range. This means the distortion may be noticed when hearing only left channel signal.

3.2. DOA information difference evaluation

Figure 4 shows the difference of DOA information between original and coded stereo signals using the proposed evaluation method. The vertical axis shows the frequency in Bark scale and the horizontal axis indicates location of sound in azimuth. The shading shows the difference of power level in each frequency and azimuth.

According to the results, as well as LD evaluation, there are no clear differences between 128 kbps and 96 kbps. However, for 80 kbps and 64 kbps, the change of DOA information is shown in lower frequency in bark scale.

4. Subjective test

In order to observe the quantitative measure, the set of subjective tests are performed, and the results are compared with the estimation of the proposed methods. Signals and coding conditions are the same as the ones of simulation. In addition, to compare with the results of DOA information, monophonic signals are prepared to inspect that how subjects can discriminate the degradation of signals only from the distortion in monaural signal.

4.1. Test procedure

ABX discrimination test is used for subjective test. Details of the task are as follows. At first, three stimuli, A, B and X are presented to subject where stimulus X is either of stimulus A or B. Subject is asked to answer which of A and B is similar to X. The task is performed 8 times to each subject in one session. Eleven subjects who have...
normal hearing ability are involved.

4.2. Results of subjective test and examination

Figure 5 shows the results of subjective test. The vertical axis shows the percentage of correct answers and the horizontal axis indicates compression bit rate. According to the results, percentages of correct answers of both 80 kbps and 64 kbps compressed signal are higher than 75% discrimination threshold. This means subjects can distinguish the difference between coded and original signals. On the other hand, percentages of correct answers for 128 kbps and 96 kbps compressed signals are about 50% to 60%. This means that subject can’t discriminate the difference. In addition, there are statistically significant differences between results of monophonic and stereo signals at 96 kbps.

Evaluation values shown in Fig.3 and Fig.4 are compared with the results of subjective test. Detectability of low bit rate signals is corresponding to the evaluation values not only due to the distortion at higher frequency but also the change of DOA information in lower frequency range. As shown in Fig.4(c) and (d), DOA information difference of 80 kbps is larger than one of 64 kbps signal. The phenomenon can be understood as a coding scheme by MPEG for lower bit rate; higher frequency bands are often omitted to assign bit rate. Although there are no apparent differences in DOA information when evaluating 96 kbps signals, the linear distortion rises detectability shown in subjective test results of monophonic signal. when the compression bit rate is set to 96 kbps. This phenomenon can be considered that binaural masking effects[5] rises masking threshold in each channel.

5. Conclusions

In this paper, an evaluation method based on a psychoacoustic model for stereo signal is proposed using frequency domain binaural model. According to the results of simulations and subjective tests, the proposed evaluation method shows the difference between original and compressed signals by means of the linear distortion and the DOA information difference. Although the relationship between evaluation by proposed method and subjective tests are shown, quantitative evaluation is necessary as future work. Also the evaluation method needs to take binaural masking effects into consideration according to the difference between spectrograms of evaluation values and results of subjective tests.

6. References