2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

심리음향모델에 근거한 잡음 형상화

이진걸
배재대학교 전자공학과

Noise Shaping Based on Psychoacoustic Model

Jingeol Lee
Department of Electronic Engineering, Paichai University

E-mail: jingeol@mail.paichai.ac.kr

Abstract - A psychoacoustic model based noise shaping method is proposed, where noise's presence with a host signal will not be perceptually noticeable. The derivation of imperceptible noise levels from the masking thresholds of the signal involves a deconvolution associated with the spreading function in the psychoacoustic model, which results in an ill-conditioned problem. In this paper, the problem is formulated as a constrained optimization, and it is demonstrated that the solution provides noise shaping where the noise excitation level conforms to the masking thresholds of the signal.

Key Words - Noise Shaping, Psychoacoustic Model, Masking, Optimization

I. INTRODUCTION

Noise shaping techniques have been widely employed in many different applications such as the perceptual weighting filter in Code-Excited Linear Predictive Coding (CELP), audio coding based on psychoacoustic model, and watermarks or data hidings for audio signals. Many of these applications exploit the masking properties of the auditory system. In order to be masked, the excitation levels of the target must fall below the masking thresholds of the masker in the auditory domain [1], [2]. The derivations of imperceptible noise levels from the masking thresholds of a host signal involve a deconvolution associated with the spreading function in the psychoacoustic model, which has been known as an ill-conditioned problem. It often leads to artifacts such as a negative energy for a threshold, zero thresholds, etc [1], [3]. In this paper, the inaudible noise levels are derived such that their excitation levels conform to the masking threshold of the host signal by formulating the problem as a constrained optimization.

II. DERIVATION OF INAUDIBLE NOISE LEVEL

The psychoacoustic model 1 employed in MPEG audio is exploited. Since the excitation level of each masker is higher than its masking threshold by the masking index, the excitation level is derived from the global masking thresholds defined in the psychoacoustic model by removing the masking index and the absolute threshold terms [4],

$$ET_g(i)=10\log\left(\sum_{j=1}^{m}10^{ET_{tm}\{z(j),\,z(i)\}/10}+\sum_{j=1}^{n}10^{ET_{nm}\{z(j),\,z(i)\}/10}\right) \qquad (1)$$

where

$$ET_{tm}\{z(j),\,z(i)\}=X_{tm}\{z(j)\}+vf\{z(j),\,z(i)\}\;dB$$
$$ET_{nm}\{z(j),\,z(i)\}=X_{nm}\{z(j)\}+vf\{z(j),\,z(i)\}\;dB.$$

$ET_g(i)$ is the global excitation level. $ET_{tm}\{z(j), z(i)\}$ and $ET_{nm}\{z(j), z(i)\}$ are individual excitation levels of the tonal and the non-tonal components, respectively, at the subsampled frequency index i corresponding to the critical band rate of $z(i)$ caused by the masking component at the index j corresponding to the rate of $z(j)$. The function vf is the spreading amount.

In order to shape noise such that its excitation level falls below the masking threshold of the host signal, it is proposed to modify the sound pressure level of the noises in each critical band, which leads to a linear filtering of the noise. Since the powers of the spectral lines are summed within each critical band to form the sound pressure level, the filter gain is assumed to be constant within each critical band so that the shaped noise is given by

$$\hat{N}(k)=H(b)N(k), \quad k_{lb}\le k\le k_{hb}, \quad 0\le b\le B-1 \qquad (2)$$

where k_{lb}, k_{hb} are the lower and upper bounds, respectively, of the critical band b, and B is the total number of the critical bands. Considering that the excitation level of the noise at a certain frequency is found by summing the spreaded sound pressure levels of neighboring tonal and non-tonal components, the noise level at that frequency can be shaped by weighting the neighboring tonal and non-tonal components as

$$\begin{aligned}\sum_{j=0}^{B-1}H^2\{b(j)\}10^{ET[b\{z(j)\},\,z(i)]/10}&=10^{ET_g(i)/10}, \quad ET_g(i)<LT_g(i)\\ &=10^{LT_g(i)/10}, \quad ET_g(i)\ge LT_g(i)\end{aligned} \qquad (3)$$

where $ET[b\{z(j)\}, z(i)]$ is the spreaded sound pressure level at the frequency index i from the individual masker of both tonal and non-tonal components at the critical band index j. $LT_g(i)$ is the global masking threshold. The filter coefficients are squared to be consistent with the power terms in (3). In case that the excitation level of the noise at a frequency index i exceeds the masking threshold of signal, the excitation level is adjusted to the masking threshold, to the level that conforms to the masking threshold. Otherwise, the excitation level is left intact. Evaluation of (3) results in the linear algebraic equation in the form $E\cdot\mathbf{h}=\mathbf{m}$. E is a matrix whose size is the number of frequencies evaluated by the number of critical bands, and elements of each row are spreaded sound pressure levels at a frequency from individual masker at the corresponding critical band. The vector $\mathbf{h}$ consists of squares of the filter coefficients, $H^2\{b(j)\}$, whose size is exactly the number of critical bands. The vector $\mathbf{m}$ consists of either the excitation levels or the masking thresholds at the frequencies of evaluation depending on their relative amplitudes as in (3), whose size is exactly the number of frequencies evaluated. The squares of filter coefficients can be solved by the method based on singular value decomposition (SVD). However, it is found that the SVD based solution occasionally results in negative values as squares of the filter coefficients depending on noise level, which gives rises to negative powers. In order to cope with this problem, the noise shaping is formulated as a constrained optimization as follows:

$$\min_{\mathbf{h}} \|E\cdot\mathbf{h}-\mathbf{m}\|_2 \quad \textit{such that} \quad 0 \le h_0, h_1, \dots, h_{B-1} \le 1 \qquad (4)$$

III. EXPERIMENTAL RESULTS

The proposed noise shaping method is tested with artificially generated signals of the sampling rate of 8 kHz. In the test, the psychoacoustic model, originally developed for MPEG audio coding, is modified to accommodate the test signals. A sinusoidal signal with the amplitude of 1000 and the frequency of 1000 Hz as the masker and a pseudorandom noise in the range of -50 to 50 as the target are generated for the test of the proposed noise shaping method. The noise is shaped by (4) using the masking thresholds of the sinusoidal signal and the excitation levels of the noise. Fig. 1 shows the masking thresholds of the sinusoidal signal, excitation levels of the input and the shaped noise, respectively. It can be seen that excitation levels of the input noise exceed the masking thresholds of the sinusoidal signal at most of regions of frequencies whereas those of the shaped noise are constrained to the masking thresholds, by which the noise is made inaudible. There are some errors in the resulting excitation levels of the shaped noise at the frequency regions where the excitation levels of the input noise exceed the masking thresholds of the sinusoidal signal. These errors are caused by the spreading function whose functional shape depends on the sound pressure level of the individual masker and by the optimization process.

IV. CONCLUSION

A psychoacoustic model based noise shaping method is proposed, where noise's presence with a host signal will not be perceptible. The inaudible noise level is derived by formulating the problem of noise shaping as a constrained optimization, by which the excitation levels of the noise result in conforming to the masking thresholds of the host signal. It is demonstrated with artificially generated signals that the proposed method shapes the noise as constrained.

ACKNOWLEDGEMENT

The author wishes to acknowledge the financial support of the Korea Research Foundation made in the program year of 1998.

REFERENCES

[1] Raymond N. J. Veldhuis, "Bit Rates in Audio Source Coding," IEEE Journal on Selected Area in Communications, Vol. 10, No.1, pp. 86-96, Jan. 1992

[2] M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear," J. Acoust. Soc. Am, vol. 66, No. 6, pp. 1647-1652, Dec. 1979

[3] James D. Johnston, "Transform Coding of Audio Signals Using Perceptual Noise Criteria," IEEE Journal on Selected Areas in Communications, Vol. 6, No. 2, pp. 314-323, Feb. 1988

[4] ISO/IEC 11172-3, "Information technology-Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, Part3: Audio"

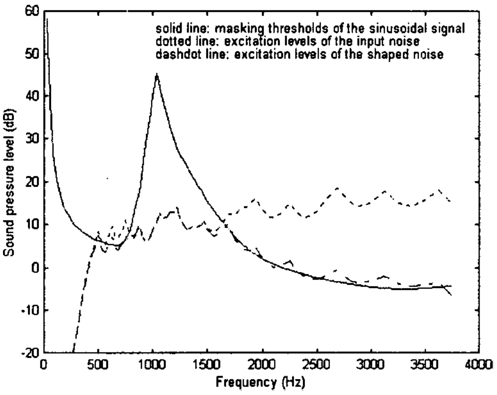

Fig. 1. Masking thresholds and excitation levels

——	Masking thresholds of the sinusoidal signal
....	Excitation levels of the input noise
—·—·	Excitation levels of the shaped noise