THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA Vol.20, No.2E 2001. 6. pp. 9~16

# Noise Shaping Based on Psychoacoustic Model

Jingeol Lee*, Seung Hyon Nam*

*Division of Information Communication Engineering, Paichai University
(Received 29 May 2000 ; revised 15 January 2001 ; accepted 12 April 2001)

**Abstract**

A psychoacoustic model based noise shaping method which shapes the noise in the frequency domain is proposed, where its 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, and thus the noises embedded in the signal will not be perceived by human ear.

***Keywords:*** *Noise shaping, Psychoacoustic model, Masking, Optimization*

## I. Introduction

Noise shaping techniques have been widely employed in many different applications in various forms. For instance, the perceptual weighting filter in Code-Excited Linear Predictive Coding (CELP) shapes the error between the original and the synthesized signal such that the frequency regions corresponding to the formants are de-emphasized [1]. The noise (error) components in the frequency region of formant nulls, where the components are more subjectively disturbing, are reduced further. Another instance is found in audio coding based on psychoacoustic model that shapes the quantization noise such that its presence is not perceptible[2]. Furthermore, in noise suppression, speech and audio enhancement applications, noise shaping techniques are involved in a sense because the frequency contents of noisy signals are modified in accordance with the frequency domain representation of estimated clean signals[3,7]. Another example includes watermarks or data hidings for audio signals which is the process to embed data such as copyright information into audio with a minimum amount of perceivable degradation to the host signal[8].

Many of these applications exploit the masking properties of the auditory system. The masking property is the psychoacoustic phenomenon that a stronger signal, e.g., a pure tone makes a weak signal, e.g., quantization noise inaudible (masked). The masking signal is called the masker; the masked signal is called the target. In order to be masked, the excitation levels of the target must fall below the masking thresholds of the masker in the auditory domain[9,10]. 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[9,11]. To get around this problem, a

Corresponding author: Jingeol Lee (jingeol@mail.paichai.ac.kr)
Paichai University, Taejon 302-735, Korea

suboptimal renormalization can be used in place of the deconvolution[11].

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. The previous work in [9] approached the problem of minimal bit rates in audio coding similarly by formulating the detection criterion of the target as constraints. The criterion adopted in this paper is that excitation levels of the noises fall below the masking thresholds of the host signal in the auditory domain while one in [9] is that excitation level of masker and target together differs by less than 1 dB from the excitation level of masker alone. It is well known in psychoacoustics that a target can be made inaudible (masked) by a masker as long as the excitation level of the target fall below the masking thresholds of the masker in the auditory domain[10-13]. From this point of view, the detection criterion employed in this paper is more generally acceptable.

The psychoacoustic model is reviewed in Section II. In Section III, the proposed method of derivation of inaudible noise levels by the constrained optimization is formulated. In Section IV, its applicabilities are demonstrated by experiments using artificial and speech signals. The conclusions are drawn in Section V.

# II. Review of the psychoacoustic model

## 2.1. Masking

The masking is a perceptual property of the human auditory system, by which the presence of a strong signal makes a spectral and temporal neighborhood of weaker signals imperceptible[14]. They are referred to as simultaneous and nonsimultaneous masking, respectively. The simultaneous masking depends on the frequency, power, and tone-like or noise-like characteristics of both the masker and target. The nonsimultaneous masking includes pre- and post-masking[8]. The pre-masking is a property that a weaker signal, existing for 5-20 msec before the strong signal, is made inaudible. The post-masking refers to a property that a weaker signal, existing for 50-200 msec after the strong signal, is made inaudible. The masking threshold associated with the post-masking is modeled as an exponential decaying function[5,12]. In this paper, the simultaneous masking model employed in MPEG audio psychoacousic model 1 for layer II is exploited.

Empirical results also show that the human auditory system has limited frequency-dependent resolution over which the human ear seems to integrate. This dependency can be expressed in terms of critical bandwidth of 100 Hz for frequencies below 500 Hz and approximately a third octave for frequencies above 500 Hz. The critical band scale is also used as a measure of frequency, called critical band rate, $z$. The corresponding unit is the Bark. Table 1 lists cutoff frequencies of the critical bands[9,14].

Table 1. Critical band frequencies, taken from [9].

| Rate (Bark) | $\Delta f$ (Hz) | $f_l$ (Hz) | $f_u$ (Hz) |
|---|---|---|---|
| 1 | - | - | 100 |
| 2 | 100 | 100 | 200 |
| 3 | 100 | 200 | 300 |
| 4 | 100 | 300 | 400 |
| 5 | 110 | 400 | 510 |
| 6 | 120 | 510 | 630 |
| 7 | 140 | 630 | 770 |
| 8 | 150 | 770 | 920 |
| 9 | 160 | 920 | 1080 |
| 10 | 190 | 1080 | 1270 |
| 11 | 210 | 1270 | 1480 |
| 12 | 240 | 1480 | 1720 |
| 13 | 280 | 1720 | 2000 |
| 14 | 320 | 2000 | 2320 |
| 15 | 380 | 2320 | 2700 |
| 16 | 450 | 2700 | 3150 |
| 17 | 550 | 3150 | 3700 |
| 18 | 700 | 3700 | 4400 |
| 19 | 900 | 4400 | 5300 |
| 20 | 1100 | 5300 | 6400 |
| 21 | 1300 | 6400 | 7700 |
| 22 | 1800 | 7700 | 9500 |
| 23 | 2500 | 9500 | 12000 |
| 24 | 3500 | 12000 | 15500 |
| 25 | 6550 | 15500 | 19500 |
| 26 | 6550 | 19500 | 24600 |

## 2.2. Review of the masking model in MPEG audio

The psychoacoustic model 1 in MPEG audio provides signal-to-mask ratios of 32 subbands, which are used in bit allocation process for the subbands. Since this paper deals with the noise shaping exploiting the simultaneous masking model in MPEG audio, the contents relevant to masking (MPEG audio psychoacousic model 1 for layer II supporting the sampling rate of 32 kHz) are summarized[2].

### 2.2.1. Calculation of the FFT for time to frequency conversion

The 1024-point FFT is taken directly to the input signal, $x(n)$, windowed by a Hann window, $h(n)$. The power density spectrum of the signal, $P(k)$ is calculated as

$$P(k) = 10\log\left|\frac{1}{N}\sum_{n=0}^{N-1} h(n)x(n)e^{-j(2\pi/N)nk}\right|^2 dB,$$
$$0 \le k \le N/2 \qquad (1)$$

where $N$=1024.

### 2.2.2. Determination of the threshold in quiet (absolute threshold)

The threshold in quiet, $LT_q(k)$, which is the sound pressure level defined in the frequency domain, is given in the model[2]. Humans perceive the sound's presence in quiet when its sound pressure level exceeds the threshold.

### 2.2.3. Finding of the tonal and non-tonal components of the audio signal

It is necessary to discriminate between tonal and non-tonal components from the FFT spectrum because the tonality of the masking component has an influence on the masking threshold. This step is accomplished by determining the local maxima, and then extracting tonal components and calculating the intensity of the non-tonal components within a bandwidth of a critical band. The following three operations are performed to obtain sound pressure levels of tonal and non-tonal components:

a) Labeling of local maxima

A spectral line $P(k)$ is labeled as a local maximum if

$$P(k) \ge P(k-1) \quad \text{and} \quad P(k) \ge P(k+1). \qquad (2)$$

b) Listing of tonal components and calculation of the sound pressure level

A local maximum is put in the list of tonal components if

$$P(k) - P(k+j) \ge 7\ dB, \qquad (3)$$

where $j$ is chosen according to

$$\begin{array}{ll} j = -2, 2 & \text{for } 2 < k < 63 \\ j = -3, -2, 2, 3 & \text{for } 63 \le k \le 127 \\ j = -6, \cdots, -2, 2, \cdots, 6 & \text{for } 127 \le k \le 255 \\ j = -12, \cdots, -2, 2, \cdots, 12 & \text{for } 255 \le k \le 500. \end{array}$$

If $P(k)$ is found to be a tonal component, then the sound pressure level of the tonal component is calculated as

$$P_{tm}(k) = 10\log\left\{10^{\frac{P(k-1)}{10}} + 10^{\frac{P(k)}{10}} + 10^{\frac{P(k+1)}{10}}\right\} dB. \quad (4)$$

Next, all spectral lines within the examined frequency range are set to $-\infty$dB

c) Listing of non-tonal components and calculation of the power

Within each of 24 critical bands, the power of spectral lines (remaining after the tonal components have been zeroed) are summed to form the sound pressure level of the non-tonal component, $P_{nm}(k)$ corresponding to that critical band.

### 2.2.4. Decimation of the maskers to obtain only the relevant maskers

The decimation is a procedure that is used to reduce the number of maskers that are considered for the calculation of the global masking threshold.

### 2.2.5. Calculation of the individual masking thresholds

Of the original 512 frequency domain samples as shown in Eq. (1), only a subset of the samples, indexed by $i$, are considered for the global masking threshold calculation. The number of samples in the subsampled frequency domain is 132.

The individual masking thresholds of both tonal and non-tonal components are given by the following expression:

$$LT_{tm}\{x(j), z(i)\} = P_{tm}\{z(j)\} + av_{tm}\{z(j)\} + vf\{z(j), z(i)\}\, dB$$

$$LT_{nm}\{x(j), z(i)\} = P_{nm}\{z(j)\} + av_{nm}\{z(j)\} + vf\{z(j), z(i)\}\, dB \tag{5}$$

where $LT\{z(j), z(i)\}$ is the individual masking threshold 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 power $P\{z(j)\}$ is the sound pressure level of the masking component at the frequency index $j$ corresponding to the critical band rate of $z(j)$. The function $av$ is the masking index, and the function $vf$ is the spreading amount.

### 2.2.6. Determination of the global masking threshold

The global masking threshold $LT_g(i)$ at the $i$ th frequency sample is found by summing the individual masking thresholds corresponding to the $j$ th tonal and non-tonal maskers and the threshold in quiet $LT_q(i)$

$$LT_g(i) = 10 \log\left(10^{LT_q(i)/10} + \sum_{j=1}^{m} 10^{LT_{tm}(z(j), z(i))/10} + \sum_{j=1}^{n} 10^{LT_{nm}(z(j), z(i))/10}\right) \tag{6}$$

where the $m$ is the total number of tonal maskers, and the $n$ is the total of non-tonal maskers.

## III. Derivation of inaudible noise level

According to the psychoacoustic model in Section 2.2, the excitation level of the noise can be obtained by removing the masking index and the absolute threshold terms from the masking threshold of the noise. In other words, the excitation level of each masker is higher than its masking threshold by the masking index[13]. The excitation level is derived from Eq. (6) by removing the masking index and the absolute threshold terms,

$$ET_g(i) = 10 \log\left( \sum_{j=1}^{m} 10^{ET_{tm}(z(j))/10} + \sum_{j=1}^{n} 10^{ET_{nm}(z(j), z(i))/10}\right) \tag{7}$$

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, and $ET_{tm}(z(j), z(i))$ and $ET_{nm}(z(j), z(i))$ are individual excitation levels of \the tonal and the non-tonal components, respectively.

The noise can be made inaudible if the excitation level falls below the masking threshold of the host signal. In order to shape noise in such a manner, 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 as shown in Section 2.2, the filter gain is assumed to be constant within each critical band so that the shaped noise is given by

$$\hat{N}(k) = H(b)N(k), \quad k_{lb} \le k \le k_{hb}, \quad 0 \le b \le B-1 \tag{8}$$

where $k_{lb}$, $k_{hb}$ are the lower and upper bounds (Table 1), respectively, of the critical band $b$, and $B$ is the total number of the critical bands. Eq. (8) states that the frequency content of the noise, $N(k)$ within the frequency range of each critical band is modified by the corresponding filter coefficients, $H(b)$. 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

$$\sum_{j=0}^{B-1} H^2\{b(j)\} 10^{ET[b(z(j)), z(i)]/10} = 10^{ET_g(i)/10}, \quad ET_g(i) < LT_g(i)$$
$$= 10^{LT_g(i)/10}, \quad ET_g(i) \ge LT_g(i) \tag{9}$$

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 $j$. The filter coefficients are squared to be consistent with the power terms in Eq. (9). In case that the excitation level of the noise at the frequency index $i$ exceeds the masking threshold of the 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. The frequencies at which Eq. (9) is evaluated may be the same as subsampled frequencies at which the masking thresholds are calculated, or the frequencies where the tonal and non-tonal components are located. Evaluation of Eq. (9) at either frequency sets results in the linear algebraic equation in the form $E \cdot \boldsymbol{h} = \boldsymbol{m}$. $E$ is a matrix whose size is the number of frequencies evaluated by the number of critical bands, and elements of each row are the spreaded sound pressure levels at a frequency from individual masker at the corresponding critical band. The vector $\boldsymbol{h}$ consists of squares of the filter coefficients, $H^2\{b(j)\}$, whose size is exactly the number of critical bands. The vector $\boldsymbol{m}$ consists of either the excitation levels or the masking thresholds at the frequencies of evaluation depending on their relative amplitudes as in Eq. (9), whose size is exactly the number of frequencies evaluated. The number of unknown, i.e., the squares of filter coefficients can be greater or less than the number of the equations depending on the frequency set, which 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 rise to negative powers. In order to cope with this problem, the noise shaping is formulated as a constrained optimization problem as follows:

$$\min_{\boldsymbol{h}} \|E \cdot \boldsymbol{h} - \boldsymbol{m}\|_2 \quad \textit{such that} \quad 0 \le h_0, h_1, \cdots, h_{B-1} \le 1 \tag{10}$$

It is expected that the resulting excitation levels of the shaped noise may be somewhat different from constrained masking thresholds because the spreading function depends on the sound pressure level of the individual masker. Since the higher sound pressure level is spreaded more gradually, the resulting noise levels are expected to fall somewhat below the constrained masking thresholds. However, it is found from a series of experiments as will be shown in the next section that the resulting shaped noise is made just noticeable by conforming its excitation level to the masking threshold of the host signal in the frequency regions where the excitation levels exceed the masking thresholds. A margin to the just noticeable noise level can be placed by incorporating a constant negative bias in the masking threshold in Eq. (10), which lowers all masking thresholds by its amount.

## IV. Experimental results

The proposed noise shaping method is tested with both artificially generated signals and real speech signals. In both tests, the masking model, originally developed for MPEG audio coding, is modified to accommodate the test signals. The sampling rate of the signals used in this test is 8 kHz whereas that in MPEG is 32 kHz as in Section 2.2. Accordingly, the frame size is reduced from 1024 samples to 256 samples to maintain the same frequency resolution of the FFT. The number of subsampled frequencies is reduced from 132 to 78, and the number of the critical bands is reduced from 24 to 17, which covers baseband of the test signals, 0 - 4 kHz. The frequency range associated with the finding tonal components of Eq. (3) is modified in accordance with the down sampling so that a local maximum is put in the list of tonal components if

$$P(k) - P(k+j) \ge 7\ dB \tag{11}$$

where $j$ is chosen according to

$$\begin{aligned} j &= -2, 2 && \text{for } 2 < k < 63 \\ j &= -3, -2, 2, 3 && \text{for } 63 \le k < 125. \end{aligned}$$

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 Eq. (10) using the masking thresholds of the sinusoidal signal and the excitation levels of the noise, evaluated at the 78 subsampled frequencies. The

Table 2. Solutions for the elements of the $h$.

| | | | |
|---|---|---|---|
| $h_0$ | 1.000000 | $h_9$ | 1.000000 |
| $h_1$ | 1.000000 | $h_{10}$ | 1.000000 |
| $h_2$ | 1.000000 | $h_{11}$ | 0.991342 |
| $h_3$ | 1.000000 | $h_{12}$ | 0.065329 |
| $h_4$ | 0.816476 | $h_{13}$ | 0.043398 |
| $h_5$ | 0.540370 | $h_{14}$ | 0.010226 |
| $h_6$ | 0.420937 | $h_{15}$ | 0.008272 |
| $h_7$ | 1.000000 | $h_{16}$ | 0.007491 |
| $h_8$ | 1.000000 | | |

solution is presented in Table 2 with the initial values of elements of the vector $h$ with all 1s, which corresponds to unshaped noise.

The subscripts of $h$ in Table 2 refers to the index of critical band as shown in Table 1. That is, $h_0$ covers the frequency region up to 100 Hz, and $h_1$ covers 100 - 200 Hz, etc. The excitation levels of the shaped noise are calculated using frequency contents of the shaped noise by Eq. (8). 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.

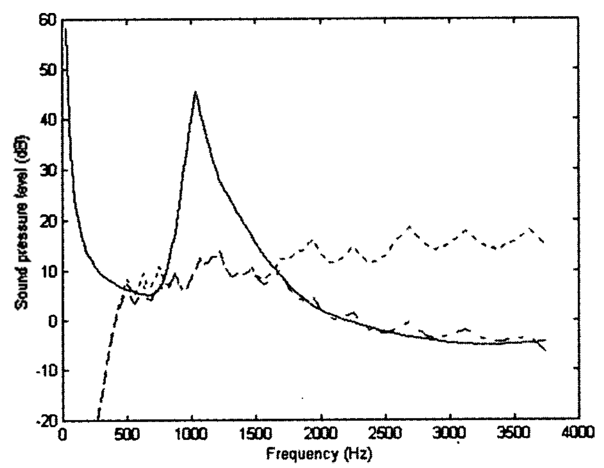



Figure 1. Masking thresholds and excitation levels.

It is noticeable that the masking thresholds and the excitation levels are spreaded across critical bands by the spreading function, and the absolute thresholds decrease within the bandwidth of 4 kHz. In addition, the excitation levels of input noise increase with frequency, caused by summing the powers of the spectral lines of broader bandwidth of the critical bands. A cross-check of the solution shown in Table 2 with the critical band frequencies of Table 1 and the resulting excitation levels of the shaped noise in Fig. 1 reveals that there are no maskers up to 400 Hz, and the frequency components of the input noise are almost left intact in the frequency regions of 770-1720 Hz while those in the regions of 400-770 and 1720-3700 Hz are attenuated as constrained. 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. A constant negative bias can be incorporated in the masking threshold in Eq. (10) in consideration of these errors or a margin to inaudibility of the noise. This test demonstrates that the proposed method shapes the input noise such that its excitation level conforms to the masking threshold of a host signal.

The above test is performed to demonstrate its applicability using a single frame of the artificial signals. As a second test, a subway noise is shaped such that the distortion of a male speech signal is not perceived with the shaped subway noise embedded. In this test, FFT size is set equal to the frame size of 256, by which it was shown that the temporal aliasing caused by circular convolution is negligible[4]. The analysis frames are overlapped with adjacent frames by 50 % and the shaped noise is obtained by the overlap addition method. In Fig. 2, time domain plots of the input speech, the subway noise, the shaped subway noise, and the input speech signal with the shaped subway noise embedded are shown.

It can be seen that the amplitude of the shaped noise changes with that of the input speech signal. This is caused by the fact that regions of the input speech signal with

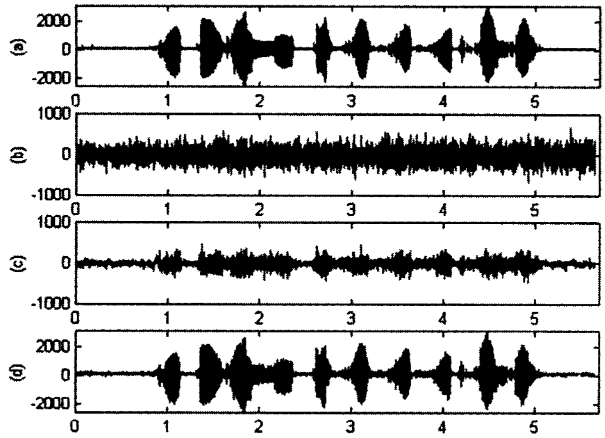



Figure 2. Waveforms.
(a) Input speech signal, (b) Input subway noise, (c) Shaped subway noise, (d) Speech signal with the shaped noise embedded

high amplitude provide high levels of the masking thresholds, which leads to high amplitude in the shaped noise. It seems that regions of silence in the shaped noise (Fig. 2 (d)) are noisier than those in the input speech (Fig. 2 (a)). However, the noises in the silence region are permitted by the absolute threshold, mostly because the regions in the input speech are almost free of noise. For subjective speech quality assessment, the pair comparisons between the input speeches and the speeches with the shaped noises embedded are given in Table 3. The pair comparisons are performed for the male speech sentence with the subway noise as in Fig. 2 and an additional female speech sentence with the bus noise. In the pair comparisons, listeners are played each pair twice and asked to choose the version they prefer.

The pair comparisons indicate that both the input speech and the speech with the shaped noise embedded are perceived as almost identical. The pair comparisons are performed with the amplitude of the signals adjusted such that their excitation levels and masking thresholds are consistent with the corresponding sound pressure levels.

## V. Conclusions

A psychoacoustic model based noise shaping method, which shapes noises in the frequency domain such that their presence with a host signal are not perceptible, is proposed. 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. Also, it is confirmed with pair comparisons that the input speech signal and the shaped noise embedded speech signal are perceived as almost identical.

The proposed method exploits the simultaneous masking property of the human auditory system. It is expected that more noise can be embedded if the nonsimultaneous masking model is incorporated along with the simultaneous masking model because the noise allowed mostly by the post-masking can be additionally embedded. Other possibility for embeding more noise is to shape the noise with an amplitude comparable to that of the host signal. The combination of these possibilities is believed to lead to the maximum inaudible noise level that can be embedded in the host signal. Therefore, the proposed method is expected to be applicable to audio watermarking robust for various manipulations and processings on the host audio signals.

Table 3. Pair comparisons.

| Male Speech Sentence with Subway Noise | | |
|---|---|---|
| Version | Preference (%) | Not Sure (%) |
| Clean Input | 5.0 | 95.0 |
| Clean Input with the Shaped Noise Embedded | 0.0 | |

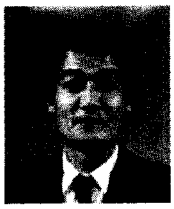
| Female Speech Sentence with Bus Noise | | |
|---|---|---|
| Version | Preference (%) | Not Sure (%) |
| Clean Input | 6.5 | 85.0 |
| Clean Input with the Shaped Noise Embedded | 8.5 | |

## References

1. A. M. Kondoz, "Digital Speech Coding for Low Bit Rate Communications Systems," John Wiley & Sons Ltd, pp. 146, 1994.
2. 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".
3. Jae S. Lim, Alan V. Oppenheim, "Enhancement and Band-width Compression of Noisy Speech," *Proceedings of the IEEE*, Vol. **67**, No. 12, pp. 1586-1604, Dec. 1979.
4. Steven F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. **ASSP-27**, No. 2, pp. 113-120, Apr. 1979.
5. Dionysis E. Tsoukalas, John Mourjopoulos, and George Kokkinakis, "Perceptual Filters for Audio Signal Enhance-ment," *J. Audio Eng. Soc.*, Vol. **45**, No. 1/2, pp. 22-36, Jan./Feb. 1997.
6. Dionysis E. Tsoukalas, John N. Mourjopoulos, and George Kokkinakis, "Speech Enhancement Based on Audible Noise Suppression," *IEEE Transactions on Speech and Audio Processing*, Vol. **5**, No. 6, pp. 497-514, Nov. 1997.
7. Stefan Gustafsson, Peter Jax and Peter Vary, "A Novel Psychoacoustically motivated Audio Enhancement Algorithm Preserving Background Noise Characteristics," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. **1**, pp. 397-400, May 1998.
8. Laurence Boney, Ahmed H. Tewfik, and Khaled N. Hamdy, "Digital Watermarks for Audio Signals," 1996 IEEE International Conference on Multimedia Computing and Systems, pp. 473-480, Jun. 1996.
9. Raymond N. J. Veldhuis, "Bit Rates in Audio Source Co-ding," *IEEE Journal on Selected Area in Communications*, Vol. **10**, No. 1, pp. 86-96, Jan. 1992.
10. M. R. Schroeder, B. S. Atal, and J. L. Hall, "Optimizing Digital Speech Coders by Exploiting Masking Properties of the Human Ear," *Journal of Acoustic Society of America*, Vol. **66**, No. 6, pp. 1647-1652, Dec. 1979.
11. 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.
12. John G. Beerends and Jan A. Stemerdink, "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation," *J. Audio Eng. Soc.*, Vol. **40**, No. 12, pp. 963-978, Dec. 1992.
13. Eberhard Zwicker and U. Tilmann Zwicker, "Audio Engineering and Psychoacoustics: Matching Signals to the Final Receiver, the Human Auditory System," *J. Audio Eng. Soc.*, Vol. **39**, No. 3, pp. 115-126, Mar. 1991.
14. Davis Pan, "A Tutorial on MPEG/ Audio Compression," *IEEE Multimedia Magazine*, Vol. **2**, No. 2, pp. 60-74, Jun. 1995.


## [Profile]


• **Jingeol Lee**



Jingeol Lee received the Bachelor of Engineering degree; the Master of Science degree from Korea University in 1981 and 1985, respectively; and the Doctor of Philosophy degree from University of Florida in 1994. From 1982 to 1990, he was with Agency for Defense Development, and from 1995 to 1996, he had worked for Samsung Electronics, where he was involved in researches and developments of military electronics throughout his job experiences. Since 1997 he has been with the Division of Information Communication Engineering, Paichai University, where he is engaged in researches in voice CODEC and remote sensing.

• **Seung Hyon Nam**

Seung Hyon Nam received the B.S.E.E. degree from Sogang University, Korea, in 1980, the M.S.E.E. degree from the University of Alabama, Huntsville, in 1987, and the Ph.D. degree from Texas A&M University, College Station TX, in 1992. From 1979 and 1985, he was with Agency for Defense Development in Korea as a Research Engineer. Since 1993, he has been with Paichai University, Taejon, Korea, where he is currently an Associate Professor in the Division of Information Communication Engineering. His research interests include speech and audio processing, data compression, adaptive filter theory.