ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

AN OBJECTIVE DISTORTION ESTIMATOR FOR HEARING AIDS AND ITS APPLICATION TO NOISE REDUCTION

Mitsunori Mizumachi and Masato Akagi
School of Information Science,
Japan Advanced Institute of Science and Technology,
1-1 Asahidai, Tatsunokuchi, Nomigun, Ishikawa 923-1292, Japan.
E-mail: {mizumati, akagi}@jaist.ac.jp

10.21437/Eurospeech.1999-577

ABSTRACT

In this paper, an objective distortion estimator called auditory-oriented spectral distortion(ASD) is proposed. It is confirmed that the ASD can accurately predict the auditory perceptual distortions represented by the mean opinion score(MOS). The ASD is used as a criterion to optimize the noise reduction algorithm for instruments that need to reduce noises appealing to the ear, for example hearing aids. Experimental results say that a suitable value of the parameter should be different for the purpose to use as between improving the auditory impressions and a front-end of speech recognition systems.

1. INTRODUCTION

There is a big demand for noise reduction in proportion to make speech processing techniques practicable. It is true that hearing aids do not come into wide use because noises in daily life give very unpleasant feelings through hearing aids. However the greater part of the existent noise reduction methods aim at speech recognition systems. The authors also have proposed a method of noise reduction using a 3ch. linear microphone array as a front-end of speech recognition systems[1]. The results of computer simulations and experiments in a real environment show that this method can reduce distortions on the LPC spectral envelopes caused by various noises[1] and answer the purpose[2]. Then the authors attempt to reduce noises appealing to the ear using this noise reduction algorithm to apply this method to various instruments specially for hearing aids.

It is necessary to prepare a criterion such as an objective distortion measure that can estimate auditory perceptual distortions before reforming the noise reduction algorithm. There are some objective distortion measures introduced the knowledge of human auditory perception[3]. They are divided into telephone-band coded speech measure(300-3400 Hz) or wideband high quality audio measure(20-20000 Hz). We need an objective distortion measure for the intermediate quality speech signals(100-6000 Hz) to keep sufficient clarity, as the telephone-band coded speech measure is insufficient and the high quality audio measure is oversufficient. That is the frequency range suitable for hearing aids.

In this paper, an objective distortion estimator that is particularly aimed at additive noises and takes account of the human auditory masking phenomena is proposed, and its application to noise reduction is discussed. It is also confirmed that this distortion estimator can accurately predict the auditory perceptual distortions by some perceptual experiments.

2. AN OBJECTIVE DISTORTION ESTIMATOR

2.1 System Overview

An objective distortion estimator is constructed by introducing the human auditory masking effects into the spectral distortion(SD) that is a famous distortion measure for estimating coded speech signals, so it is called the auditory-oriented spectral distortion(ASD). Characteristics of both simultaneous and temporal masking are included in the ASD assuming that they do not depend on sound pressure levels of the target signals to simplify its calculation. We calculate the ASD using spectral components over the masked thresholds in the same way as the segmantal SD in each short term frame whose length is 21.3 msec and period is 5.3 msec.

2.2 Implementation of Simultaneous Masking Effect

The masker candidates are detected from a target signal and the masking pattern for each masker is calculated.

2.2.1 Detection of Masker Candidates

First of all, a target signal $x(t)$ passes through the A-characteristic filter adopted in sound level meters. An A-weighted amplitude spectrum $X(\omega)$ must be like to the human perception in comparison with an

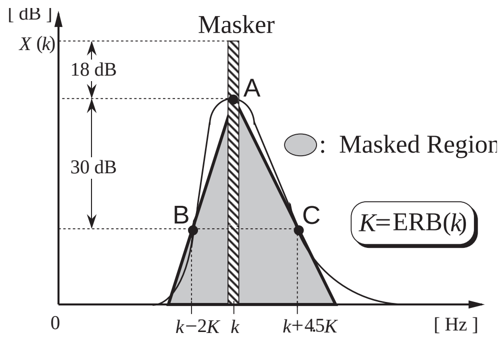

Figure 1: A masking pattern as an approximation of the masking pattern for narrow-band noises measured by Egan and Hake.

amplitude spectrum itself. When a spectral component $X(k)$ satisfies all condition in Eq. (1), it is detected as one of masker candidates.

$$\begin{cases} X(k) > X(k-1), \\ X(k) \geq X(k+1), \\ X(k) - X(k+j) > 3 \text{ [dB]},\ j = 1, 2, \cdots, J. \end{cases} \tag{1}$$

The search range J in Eq. (1) is defined as the bandwidth of an auditory filter whose center frequency is k. We reduce the number of masker candidates on the another assumption that there is a masker at the most in an auditory filter.

2.2.2 Calculation of Masking Patterns

A masking pattern for a masker candidate is calculated as an approximation of the masking pattern for narrow-band noises measured by Egan and Hake[4]. For practical purpose, the masking pattern measured in perceptual experiment is approximated by three points as follows:

$$\begin{cases} A: & (k,\ X(k) - 18), \\ B: & (k - 2 \cdot \mathrm{ERB}(k),\ X(k) - 48), \\ C: & (k + 4.5 \cdot \mathrm{ERB}(k),\ X(k) - 48), \end{cases} \tag{2}$$

where $X(k)$ means the sound pressure level in dB of a masker candidate in k Hz, and $\mathrm{ERB}(k)$ means the bandwidth in Hz of an auditory filter whose center frequency is k Hz. Each masking pattern is calculated as a triangle depended on three points in Eq. (2) such as the blacken region in Fig. 1.

2.3 Implementation of Temporal Masking Effect

The effect of temporal masking is implemented by attenuating masking thresholds calculated in the past frames. The authors adopt the post-masking curve that is measured by Zwicker[5] and shown in Fig. 2 to decide the elapsed attenuation level in dB of a masker candidate. In this paper, the effect of backward masking is disregarded because it is not far and away influential than the effect of forward masking.

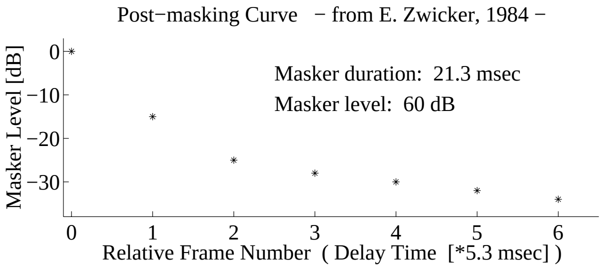

Figure 2: A post-masking curve measured by Zwicker.

2.4 Calculation of the ASD

We can obtain a masking threshold in each frame by integrating all masking patterns for simultaneous and temporal masking using the power-law model[6]. For the masked threshold $X_{\omega_i}(\omega)$ from a single masker whose frequency is ω_i, a total masked threshold $X_{total}(\omega)$ is calculated as follows:

$$X_{total}(\omega) = \mathrm{inv}F\left[\sum_{\omega_i} F\left[X_{\omega_i}(\omega)\right]\right],\ F(z) = z^p, \tag{3}$$

where p is a constant. Lutfi et al. have confirmed that $p = 0.33$ is the most desirable as compared with results of perceptual experiments[6] when only a few maskers exist unlike speech signals.

Here, assuming that we can not perceive distortions if their spectral levels are under masking thresholds, the ASD is calculated as well as the SD using only the spectra over the masking threshold.

3. VERIFICATION OF THE ASD

The authors have verified whether the ASD works well. A criterion for objective distortion measures is the linearity between subjective evaluations and objective evaluations. We can easily estimate how much distorted a speech signal is when the linearity is realized. In this paper, the authors have examined the relationship between the mean opinion score(MOS) as a subjective evaluation and the ASD or the SD in some experiments. Then the ASD differs according to the parameter p in Eq. (3) which varies from 0.3 to 0.9 in 0.1 steps. The optimization of the parameter p in Eq. 3 was also conducted.

3.1 Procedure

To obtain the MOS, a continuous Japanese vowel /ao/ in the ATR speech database is prepared and distorted by adding a random noise (from 2 kHz to 3 kHz) in various levels. The level of the random noise is determined just as the ASD or the SD in dB is 0.25, 0.5, 0.75 times as large as each maximum noise level in the ASD or the SD that is calculated against

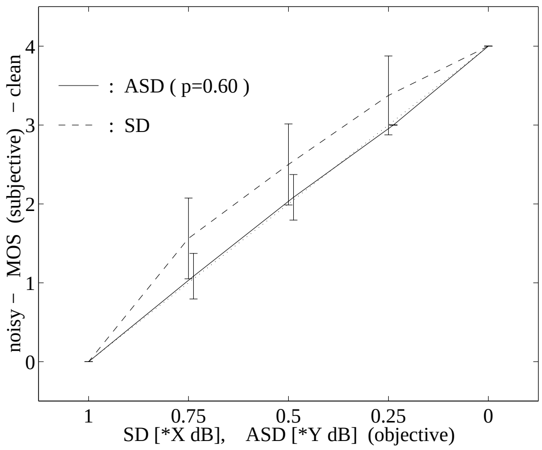

Figure 3: An interrelated curve between the ASD(a solid line) and the SD(a broken line), and the MOS.

the same speech signal. Test signals are presented to four subjects who are all trained male students through the headphone(STAX Lambda Nova Signature). Sound pressure levels of a clean speech signal and a noise maximum speech signal are 66 dB(A) and 75 dB(A) respectively. Subjects listen a clean speech, a noise maximum speech, and a speech to evaluate in order, or a noise maximum speech, a clean speech, and a speech to evaluate in order, and give the MOS on the five point scale to the third speech.

3.2 Experimental Results

The interrelated curves that display the relationships between the MOS and the ASD or the SD were plotted. For example, an interrelated curve that we set a parameter p in Eq. (3) by 0.60, is shown in Fig. 3. Fig. 3 shows the means of the MOS on each objective distortion measure using a solid line in the ASD and a broken line in the SD, and the standard deviations. The authors also confirmed that a scattering of the MOS among subjects was very small.

If an objective distortion estimator can perfectly predict the MOS, the interrelated curve becomes a straight line called the best desirable line(BDL). Another index to valuate the distortion measure is the cross-correlation coefficient between an interrelated curve and the BDL. In this experiment, those of the ASD(p = 0.60) and the SD were 0.992 and 0.951 respectively. Fig. 4 shows correlation coefficients between the interrelated curve for the ASD in each p and the BDL.

3.3 Discussion

Results of perceptual experiments suggest that the parameter p should be set as 0.60 in Fig. 4. Other

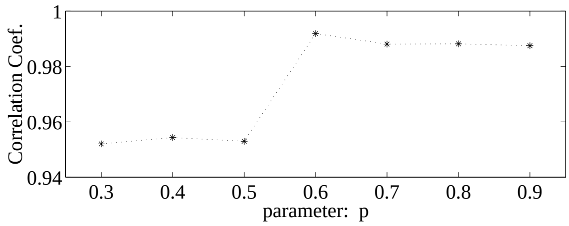

Figure 4: Correlation coefficients between the interrelated curve for the ASD on each p and the BDL.

groups also use the power-law model in the objective distortion estimator called the PAQM[3] that is based on physiological mechanisms and aims at coded speech signals whose qualities are higher than those dealt with in this paper. They decided parameter p as 0.60 or 0.80 by a lot of large-scale experiments. Although mechanisms of the ASD and the PAQM are not always same, the suitable value of parameter p for speech signals is close because both models have aimed to simulate the human auditory system. It is obvious from Fig. 3 and Fig. 4 that the ASD(p = 0.60) compared with the SD is extremely well suited to the MOS. Furthermore results of t-test show that there is a distinct difference between the ASD and the SD on condition that each standard deviation is alike.

4. APPLICATION TO NOISE REDUCTION

In order to apply the noise reduction method[1] to instruments that need to reduce noises appealing to the ear, for example hearing aids, we reconsider its parameter using the ASD.

4.1 Outline of Noise Reduction Algorithm

This noise reduction method consists of three modules: estimation of a speech and a noise directions, estimation of the noise spectrum, and subtraction of it using the Spectral Subtraction(SS)[7]. Since this method can analytically estimate noise spectra frame by frame, it is easy to reduce non stationary noises that can not be reduced by the conventional SS methods with small computational costs. In this paper, we have an eye on the SS module that is described as follows, if an amplitude spectrum of a received signal is $C(\omega)$, that of an estimated noise signal is $\widehat{N}(\omega)$ and that of a noise-reduced signal is $\widehat{S}(\omega)$.

$$|\widehat{S}(\omega)| = \begin{cases} |C(\omega)| - \alpha \cdot |\widehat{N}(\omega)|, & |C(\omega)| \geq \alpha \cdot |\widehat{N}(\omega)|, \\ \beta \cdot |C(\omega)|, & \text{otherwise}, \end{cases} \tag{4}$$

where α is the subtraction coefficient and β is the flooring coefficient. These parameters have been set

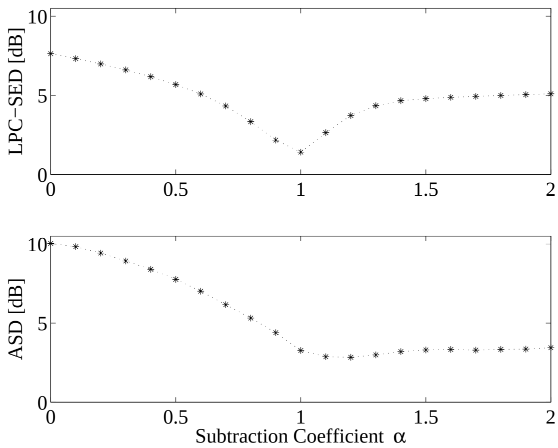

Figure 5: Distortions on the LPC-SED and the ASD of noise-reduced speech signals for each subtraction coefficient α.

as 1.0 and 0.001 respectively as a front-end of speech recognition systems.

4.2 Optimization of the Subtraction Coefficient α

The authors have tried to optimize the subtraction coefficient α based on the optimized ASD($p = 0.60$) as this is more influential than the flooring coefficient β. Test signal is the same as used in the experiment to verify of the ASD. In Fig. 5, we can see the relationships between the parameter α and distortions after noise reduction on the LPC-SED that is compatible with speech recognition rates[2] and the ASD($p = 0.60$). Note that noise reduction does not work in case $\alpha = 0$.

4.3 Discussion

Experimental results say that the most suitable value of α is different in each objective distortion measure. To improve the auditory impressions as much as possible, it is desirable that we set the parameter α by a few larger value than 1.0 that is the most suitable as a front-end of speech recognition systems. In informal tests to evaluate the qualities of noise-reduced signals by actually listening, it was confirmed that the suitable range of α is between 1.0 and 1.5 as well as the result of optimization based on the ASD($p = 0.60$). We can reconfirm that the ASD can accurately estimate distortions on auditory impressions.

On the analogy that the most suitable value of α based on the ASD is larger than 1.0, the human auditory system is sensitive to noises and insensitive to a little lacks of spectra of speech signals. In other words, spectral peaks is dominant and spectral dips is not influential for speech perception. We should take advantage of the human auditory system to construct an efficient algorithm to reduce noises appealing to the ear.

5. CONCLUSION

This paper proposed an objective distortion estimator called the ASD that takes account of auditory masking effects. It is confirmed that the ASD can accurately estimate distortions on auditory impressions measured as the MOS. Then the subtraction coefficient α used in the noise reduction method proposed by the authors is optimized using the ASD. Experimental results say that the suitable value of α is different to improve the auditory impressions as compared with the use as a front-end of speech recognition systems. It is glad that we have obtained a useful distortion estimator and it must help to equip hearing aids with noise reduction.

ACKNOWLEDGMENT

The authors would like to thank K. Kawamoto for helpful suggestions about hearing aids. This work is partly supported by CREST project under the sponsorship of Japan Science and Technology Corporation, and Research Grant from the Ministry of Education, Science and Culture of Japan(No. 10680374).

REFERENCES

[1] Mizumachi M. and Akagi M., "Noise reduction by paired-microphones using spectral subtraction," Proc. ICASSP'98, SP 30.3, Seattle, USA (1998).

[2] Mizumachi M. and Akagi M., "Noise reduction method that is equipped for a robust direction finder in adverse environments," Proc. Workshop on Robust Methods for Speech Recognition in Adverse Conditions, Tampere, FINLAND (1999).

[3] Kahrs M. and Brandenburg K., "Applications of digital signal processing to audio and acoustics," Kluwer Academic Publishers (1998).

[4] Egan J. P. and Hake H. W., "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, pp. 622-630 (1950).

[5] Zwicker E., "Dependence of post-masking on masker duration and its relation to temporal effects in loudness," J. Acoust. Soc. Am. **75**(1), pp. 219-223 (1984).

[6] Lutfi R. A., "Additivity of simultaneous masking," J. Acoust. Soc. Am. **73**(1), pp. 262-267 (1983).

[7] Boll S. F., "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. ASSP, vol. 27, no. 2, pp. 113-120 (1979).