社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report
SP2005-172 (2006-03)

Speech Processing for Hearing-Impaired Listeners Considering Threshold Elevation in the Critical Band with an Expanded Auditory Filter

Shinji MINAMIHATA †, Keiichi YASU †, Kei KOBAYASHI †,
Takayuki ARAI †, and Mitsuko SHINDO ‡

† Department of Electrical and Electronics Engineering, Sophia University

‡ Research Center for Communication Disorders, Sophia University

7-1 Kioi-cho, Chiyoda-ku, Tokyo, 102-8554 Japan
E-mail: † s-minami@sophia.ac.jp

Abstract Elevation of the threshold of audibility occurs in hearing-impaired people, and these individuals have an expanded auditory filter (Glasberg and Moore, 1986). Threshold elevation is assumed to occur due to an increase in frequency components that pass the auditory filter; an assumption known as the "power spectrum model" of masking (Patterson and Moore, 1986). Therefore, we attempted here to remove from the speech signal the frequency components that are not related to speech perception, but are instead related to threshold elevation. We calculated the masking pattern using the spreading function (Painter and Spanias, 1997), and processed monosyllabic speech samples using nine kinds of masking patterns. Both normal-hearing and hearing-impaired subjects evaluated the intelligibility and sound quality of the original and processed monosyllables. For hearing-impaired subjects, the intelligibility of a small number of certain processed monosyllables increased, but sound quality did not improve. For normal-hearing subjects, speech intelligibility decreased as the masking pattern expanded, and application of the proposed method showed no significant improvement in sound quality.

Keyword Auditory Filter, Threshold of Audibility, Critical Band, Spreading Function

1. INTRODUCTION

Glasberg and Moore (1986) measured the auditory filter of hearing-impaired people and normal-hearing people using the notched-noise masker method and found that those with hearing-impairment had a more expanded auditory filter than those without [1]. It is usually assumed that the threshold for a frequency component is determined by the amount of noise passing through the auditory filter. This assumption is referred to as the "power spectrum model" of masking (Patterson and Moore, 1986) [2]. In the virtual bandwidth of the auditory filter, known as the critical band [3], the energy of each frequency component of the input signal is summed, and the threshold of important frequency components related to speech perception is elevated [3]. When the threshold is higher than the frequency component, we cannot perceive the frequency component [3]. Based on the "power spectrum model" of masking, we hypothesized that the threshold of hearing-impaired people was more elevated than that of normal-hearing people, which would cause the decrease of intelligibility by hearing-impaired people. Therefore, with reference to the "power spectrum model", we attempted here to remove the frequency components from a speech signal that are not related to speech perception, but rather are related to threshold elevation. We calculated the masking pattern using the "spreading function" [4], and processed monosyllabic speech samples using a variety of masking patterns.

2. METHOD

We used the "spreading function" which compresses audio information in MPEG Audio [5]. Among the several types of MPEG Audio available, we used MPEG2 Audio Layer III. MPEG Audio treats the frequency components as a Bark scale when processing a speech signal. In the Bark scale, the masking pattern assumes almost the same shape regardless of the value of the center frequency of the critical band. Thus, the same masking pattern for each frequency component is used. The simplified model of the masking pattern is the "spreading function" [4], given in

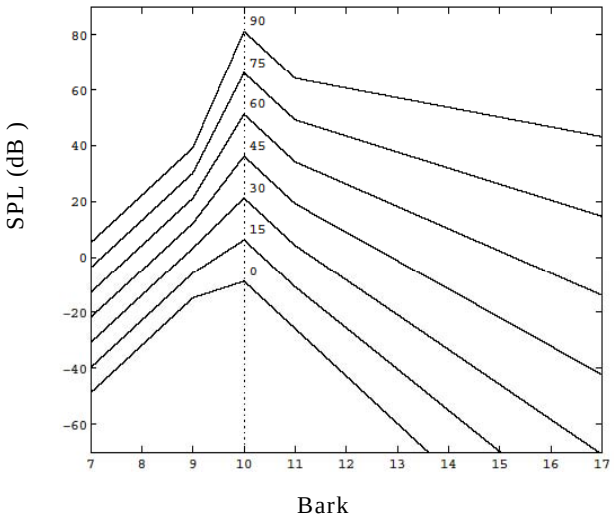

Figure 1: Spreading function (from [4]).

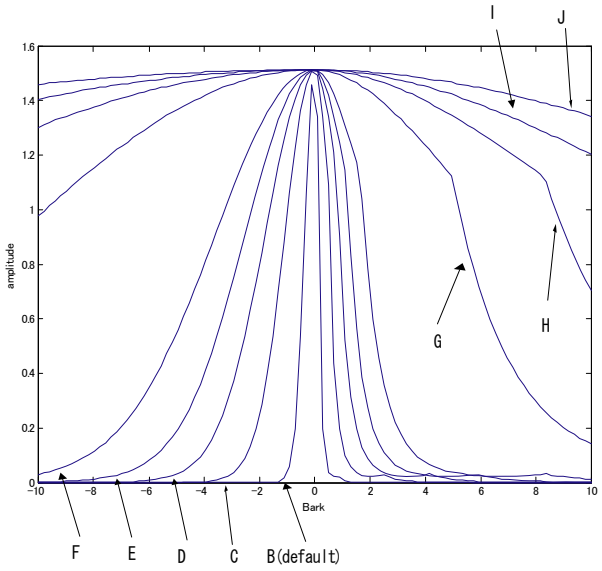

Figure 2: All 9 types of masking thresholds used.

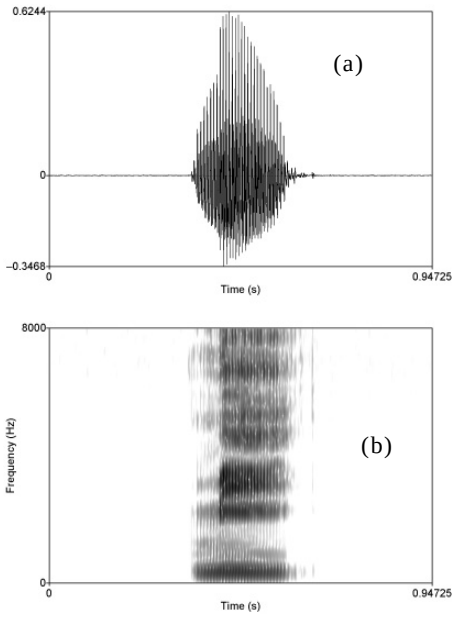

Figure 3:

Waveform (a) and spectrogram (b) of the original speech sample /mi/.

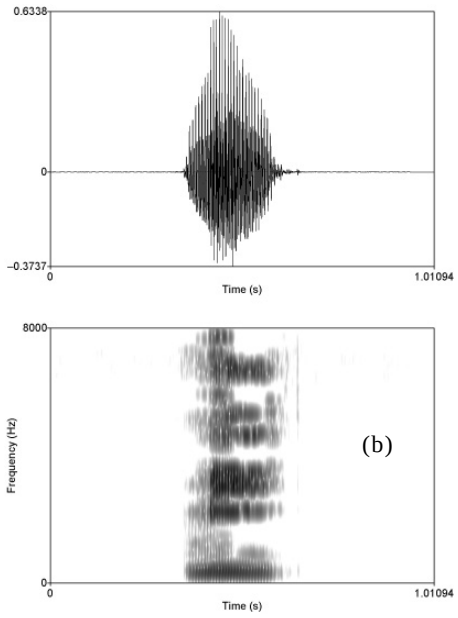

Figure 4:

Waveform (a) and spectrogram (b) of the processed signal with Type D masking threshold (/mi/).

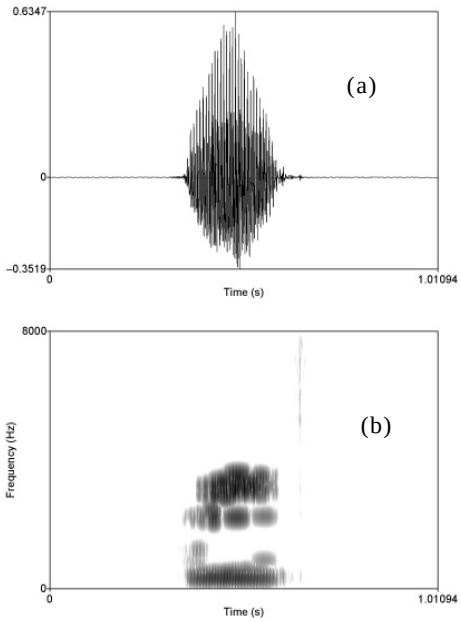

Figure 5:

Waveform (a) and spectrogram (b) of the processed signal with Type G masking threshold (/mi/).

Formula (1). Figure 1 shows the shapes of spreading function at different input intensities.

$$SF(b) = 15.81 + 7.5(b + 0.474) - 17.5\sqrt{1 + (b + 0.474)^2}\ (\mathrm{dB}) \cdots (1)$$

(b(f): Bark; SF(b): spreading function)

We expanded the shape of the masking threshold (the masking pattern for signal encoding) using the spreading function so as to remove more frequency components from the wider auditory filter of hearing-impaired people in order to improve their speech perception.

Figure 2 shows the shape of all 9 types of masking thresholds we used. Type B is the original masking threshold in MPEG2 Audio Layer III. We expanded the shape of the masking threshold to Types C to J, as shown in Figure 2.

Figure 3 shows the waveform (a) and spectrogram (b) of the original Japanese speech sample /mi/. Figures 4 and 5 show the processed signals of /mi/ using Type D and G functions, respectively.

3. Experiments

To evaluate processing, we conducted experiments with two hearing-impaired subjects (Subjects A1 and A2) and two normal-hearing subjects (Subjects B1 and B2). Both hearing-impaired subjects had hearing levels above 85 dB. All subjects evaluated the intelligibility and sound quality of the original and processed monosyllables. According to a previous study by Moore (1987), the auditory filter of normal-hearing people was expanded when they heard sounds with high sound pressure levels [6]. Therefore, by applying our processing using loud sounds, we expected that normal-hearing people would also experience some improvement.

3.1. Stimulus

We processed speech samples using Types B to J functions (9 cases). In Type A condition, the original speech samples were used without processing. Type B processing yielded the masking pattern of normal-hearing people (default setting in MPEG Audio). As speech samples, we used 24 consonant-vowel (CV) monosyllables (spoken by Japanese male) from the ATR Speech Database of Japanese (Table 1). The total number of stimuli was 240 (24 CVs × 10 processing types).

Table 1: Twenty-four nonsense consonant-vowel monosyllables (CVs) used in the experiments.

	Voiceless C+ Vowel	Voiced C+ Vowel
Stop C+ Vowel	/pa//ta//ka//pi//ki/	/ba//da//ga//bi//gi/
Fricative C+Vowel	/sa//ʃa//ha//ʃi//hi/	
Affricate C+ Vowel	/tʃa/ /tʃi/	/dza/ /dʒa//dʒi/
Nasal C+ Vowel		/ma/ /na/ /mi/ /ni/

3.2. Procedure

The experiment was controlled by a personal computer and was conducted in a soundproof room. Stimuli were presented by a loudspeaker (BOSE 402 Professional Loudspeaker System). Both hearing-impaired subjects used their own hearing aids (Subject A1: RION HB79P; Subject A2: Widex P38-VC) and the volume was set to a comfortable level for each subject before the experiment commenced. For normal-hearing subjects, the intensity of the signal was 90 dBL_A.

For each subject, 240 stimuli were randomly presented. Subjects could listen to the stimulus only once. In the open intelligibility test, the subjects were instructed to indicate which monosyllable they heard. In the sound quality test, the subject answered two closed questions: first, “Is the sound easy to hear?” and second, “Is the sound too loud?” The intelligibility test and sound quality test were conducted separately.

4. Results and Discussion

4.1. Hearing-impaired subjects

Figure 6 shows the results of the intelligibility tests for the hearing-impaired subjects (Subjects A1 and A2). Figures 7 and 8 show the results of the sound quality test for Subjects A1 and A2.

4.1.1. Subject A1

In the intelligibility test, intelligibility for Type B processing was higher than that for the original sounds (Type A). However, the subject gave only one more correct answer for Type B processing than for Type A, and thus we found no effect of our processing on intelligibility for Subject A1. The intelligibility of almost all monosyllables processed was found to be decreased as the masking threshold expanded. However, Subject A1 answered correctly for most of the processed /tʃi/ monosyllables, with the exception of the original sound (Type A). Also, for processed monosyllables /ka/, /ʃi/ and /ma/, Subject A1 showed high intelligibility.

With regard to the sound quality test, we found no positive effect related to our processing (Figure 7).

4.1.2. Subject A2

In the intelligibility test, intelligibility for Types F and H processing was higher than that for the original sounds. However, as for Subject A1, there was only one or two more correct answers on processing Types F and H than for the original sounds. Thus, intelligibility for Subject A2 was also not affected by our processing. However, for monosyllables /na/ and /ha/, though Subject A2 mistook

the original sound, he answered correctly for several processed sounds, although we could not find the relationship between the degree of expansion of the masking threshold and correct answers.

We also found no effect of our processing on sound quality for Subject A2 (Figure 8).

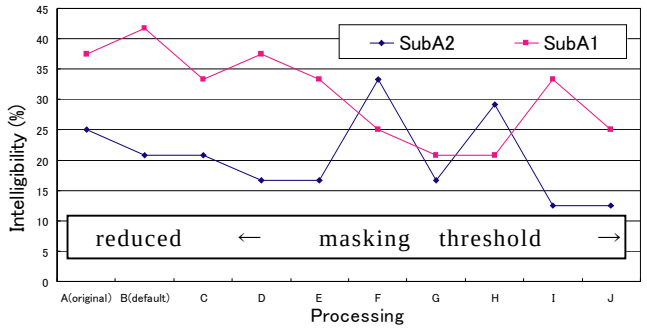

Figure 6: Results of intelligibility test for Subjects A1 and A2.

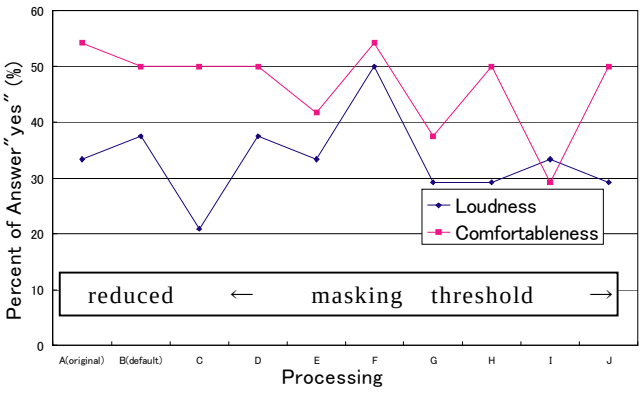

Figure 7: Result of the sound quality test for Subject A1.

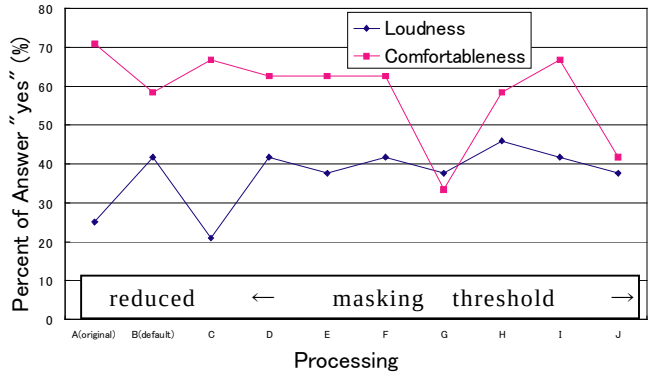

Figure 8: Result of the sound quality test for Subject A2.

4.2. Normal-hearing subjects

Figure 9 shows the results of the intelligibility tests for the two normal-hearing subjects, Subjects B1 and B2. Intelligibility of both subjects was decreased as the masking threshold expanded. This resulted from an excessive decrease in the important frequency components required for speech perception when using an expanded masking threshold.

Figures 10 and 11 show the results of the sound quality tests for Subjects B1 and B2, respectively. The evaluation of loudness was not improved by any processing. We presented the stimulus at 90 dBL_A, which is very high intensity, and it is therefore possible that the subjects could not recognize the difference in loudness of several stimuli.

In the ease-of-listening test, the evaluation score decreased as the masking threshold expanded. This result differed from that for hearing-impaired subjects. The normal-hearing subjects were able to recognize the change in ease-of-listening for the different stimuli. Musical noises occurred in many stimuli. We can observe the cause of musical noise generation as multiple isolated spots in the spectrogram [7] shown in Figure 5b. Our processing removes the frequency components whose levels are under the level of the global masking threshold. Therefore, musical noise can occur in these stimuli. This might lead to decreased ease-of-listening for normal-hearing subjects.

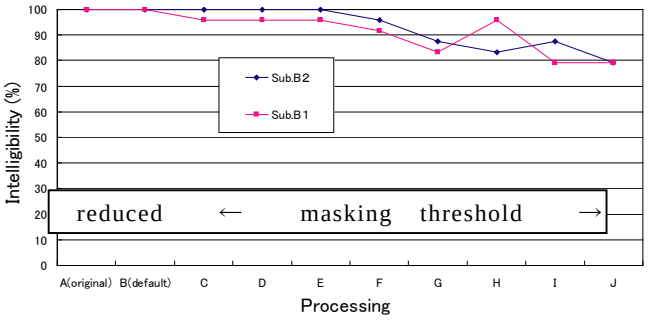

Figure 9: Results of the intelligibility tests for Subjects B1and B2.

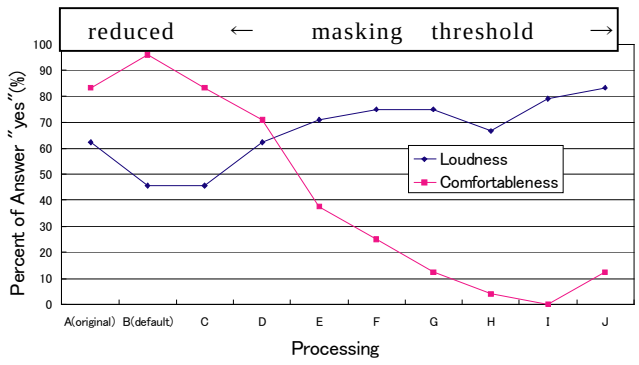

Figure 10: Result of the sound quality test for Subject B1.

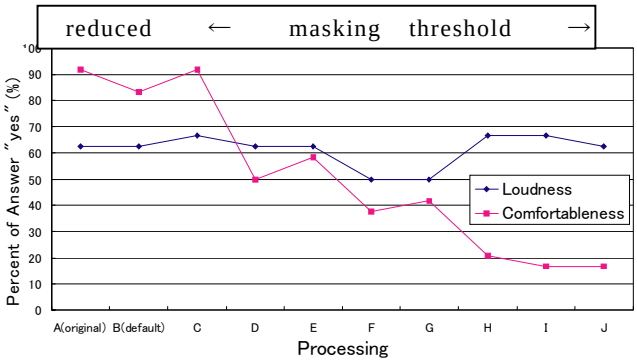

Figure 11: Result of the sound quality test for Subject B2.

5. Conclusions

We were not able to identify a suitable masking threshold for either the hearing-impaired or normal-hearing subjects. However, hearing-impaired subjects were able to correctly identify some monosyllables for certain processing types, even when they mistook the original sounds (Type A). We propose that if we could measure the shape of the auditory filter for each hearing-impaired subject, we would be able to select the most suitable masking pattern for every monosyllable. As the number of subjects was only four in the present study, we plan to conduct further experiments with a larger number of subjects in order to accurately evaluate our processing.

6. Acknowledgments

This research was supported in part by Grants-in-Aid for Scientific Research (A-2, 16203041) from the Japan Society for the Promotion of Science. We would like to thank all subjects who participate the experiments.

7. References

[1] B. R. Glasberg and B.C.J. Moore, "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. 79: 1020-1033, 1986.

[2] R. D. Patterson and B. C. J. Moore, "Auditory filters and excitation patterns as representations of frequency resolution," Frequency Selectivity in Hearing, B.C.J. Moore, ed., Academic, London, 1986.

[3] B. C. J. Moore, An Introduction to the Psychology of Hearing, 3rd edition, Academic Press,London, 1989.

[4] T. Painter, and A. Spanias, "A review of algorithms for perceptual coding of digital audio signals," International Conference on Volume 1, 179 – 208, 1997.

[5] Y. Hujiwara, "Textbook of MPEG for Practice," ASCII publication, 1995.

[6] B. C. J. Moore and B. R. Glasberg, "Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns." Hear. Res. 28, 209-225.

[7] Y. Nomura, H. Tozawa, N. Yamashita, J. Lu H. Sekiya and Takashi Yahagi, "Musical Noise Reduction by Spectral Subtraction Using Morphological Filter." 2005 RISP International Workshop on Nonlinear Circuit and Signal Processing, pp.415-418, Mar. 2005.

Appendix

In this study, we modified the “LAME” MP3 encoder to create several masking thresholds. The program shown below is the part used for calculating the masking threshold (lame-3.96.1/libmp3lame/psymodel.c). We changed parameters (1) and (2) in the source code. Table 2 shows how we changed the parameters on each processing.

```
/*
The spreading function. Values returned in units of energy
*/
static FLOAT8 s3_func(FLOAT8 bark) {
      FLOAT8 tempx,x,tempy,temp;
      tempx = bark;
      if (tempx>=0) tempx *= 3;・・・・・・( 1 )
      else tempx *=1.5;・・・・・・(2)

      if (tempx>=0.5 && tempx<=2.5)
        {
          temp = tempx - 0.5;
          x = 8.0 * (temp*temp - 2.0 * temp);
        }
      else x = 0.0;
      tempx += 0.474;
                          tempy = 15.811389 + 7.5*tempx -
                    17.5*sqrt(1.0+tempx*tempx);

      if (tempy <= -60.0) return    0.0;
      tempx = exp( (x + tempy)*LN_TO_LOG10 );

/* Normalization. The spreading function should be
normalized so that:
             +inf
               /
               |   s3 [ bark ]   d(bark)     =   1
               /
             -inf
*/
      tempx /= .6609193;
      return tempx;
}
```

Table 2: Parameters (1) and (2) on each processing.

Parameter	B(default)	C	D	E
(1)	3	3/3	3/5	3/7
(2)	1.5	1.5/3	1.5/5	1.5/7

F	G	H	I	J
3/10	3/30	3/50	3/70	3/100
1.5/10	1.5/30	1.5/50	1.5/70	1.5/100