Study of Improvement of Intelligibility for the Elderly Speech Based on Formant Frequency Shift

[1, *] Yuto Tanaka, [1]Hiroaki Igaue, [1]Mitsunori Mizumachi and [1]Yoshihisa Nakatoh

Abstract

This study aims at improving intelligibility of Japanese elderly speech based on the characteristics. In general, it is difficult for elderly people to tune aging articulators accurately. In the previous research, it was found that the transition distance of non-briskness elderly speech decreased. In addition, it was found that there was relationship between the formant and transition distance.

In this paper, we investigated the relationship of difference of transition distance and position of articulation between the briskness and the non-briskness speech of Japanese elderly speakers. In addition, we improved the sound of the elderly people by some signal processing. Specifically, the first formant frequency (F1) and the second formant frequency (F2) shift the method which makes briskness speech close to non-briskness speech.

As the results of analyzing non-briskness and briskness speech, we found that there was a relationship between the difference of transition distance and position of articulation. It is difficult for elderly speakers to utter /a/ and /i/ and to move a large mouth and tongue; thus the vowels expect /u/ converges to /u/ in F1 and F2. And the F1 and F2 shift method improved non-briskness speech intelligibility in the words which need large movements of mouth and tongue. In future works, we will investigate analysis and how to improve acoustic features of consonants in the elderly speech.

Keywords: elderly speech, formant frequency shift.

1. Introduction

Speech is very important for us to communicate with others in our daily lives. However, because of aging, some people’s ways of speech change. Owing to it, they cannot have conversation smoothly. For the aging society worldwide, this problem is very serious. For smooth communication, it is necessary to improve the elderly speech.

The previous research reported about analyzing Japanese elderly speech, and found that elderly speech have no intelligibility as compared with non-elderly by listening test [1]. And, there is a correlation between the intelligibility and the difference of transition distance [2]. However, the relationship of the utterance behavior is not investigated.

*Corresponding Author: Yuto Tanaka
(E-mail: o595401y@mail.kyutech.jp)
[1]Department of Electric and Electrical Engineering Kyushu Institute of Technology
1-1, Suisen-ch,o tobata-ku, kitakyushu city, Fukuoka 804-8550

In this study, we improved speech of elderly people based on the analysis. As analyzing, we investigated the relationship between difference of transition distance and position of articulation in terms of the level of briskness. As the method for improving, we used shift method based on formant frequency. Specifically, the first formant frequency (F1) and the second formant frequency (F2) shift method which makes briskness speech close to non-briskness speech.

2. Database of Elderly Speech

This chapter describes database of the elderly speech e and the method for selecting the subjects.

2.1 Recording

We recorded speeches of the 36 male elderly persons over the age of 60 in order to improve the intelligibility of the elderly speech. Recorded words are 543 isolated words which have phoneme balance. The elderly speech was recorded on 16-bit, 24k sampling. Table.1. shows recorded number of elderly people by age.

2.2 Subjective characteristics of elderly speech

For analyzing influence of aging intelligibility more, we selected the speaker with the impression conspicuous based on people's feeling impression of subjective characteristics of the elderly speech and analyzing the physical features of the speaker's voice.

We conducted a listening test to determine the degree of subjective characteristics of the elderly speech that include “rough”, “slow speaking”, and “non-briskness”[3]. The subjects were 10 male and 10 female students, and they were asked to judge the degree of subjective characteristics on a five-point scale from 1 to 5 where a higher score indicated a less characteristic voice. The subjects listened to 50 connected words, which were prepared from a phonetically balanced 543-word database, spoken individually by 36 elderly male speakers. Each speaker was labeled with the degree of characteristics based on the five-point scale. Each figure of the degree of characteristics of the elderly speech in "non-briskness" is carried as shown in Fig. 1. . A vertical axis expresses the evaluation degree in each impression of characteristics of the elderly speech, and a horizontal axis rearranges each evaluation speaker's value according with the degree of an evaluation.

Table 1: Number of elderly people by age

ages	60~69	70~79	80~
Number of men	16	18	2

In order to analyze about the elderly speech, we selected six speakers with non-briskness and six speakers with briskness.

3. Relationship Between Difference of Spectral Transition Distance and Position of Articulation

We noted that the degree of reduction of the transition distance is different for each vowel. We studied the relationship between the position of articulation and the difference of transition distance using non-briskness speech and briskness speech.

3.1 Spectral transition distance

A spectral transition distance is calculated by the following Eq. (1).

$$DIS = \sqrt{\sum_{n=1}^{30} \left(C_n^{(t1)} - C_n^{(t2)} \right)^2} \qquad (1)$$

$\mathbf{C_n}$ is cepstrum coefficient of order **n**. A cepstrum is the result of taking the Inverse Fourier transform (IFT) of the logarithm of the estimated spectrum of a signal. $\mathbf{t_1}$ and $\mathbf{t_2}$ are the time in the feature points of each phoneme. The feature points represent the most characteristic of each phoneme. $\mathbf{t_1}$ is the time of speech with briskness. $\mathbf{t_2}$ is the time of speech with non-briskness. In this chapter, **n** is 30 to calculate the difference between the feature quantity of speeches. A previous research shows there is relationship between a degree of briskness and spectral transition distance [3]. And, there are different features of phoneme in initiate and medial [2]. We distance on elderly speeches for each initiate and medial.

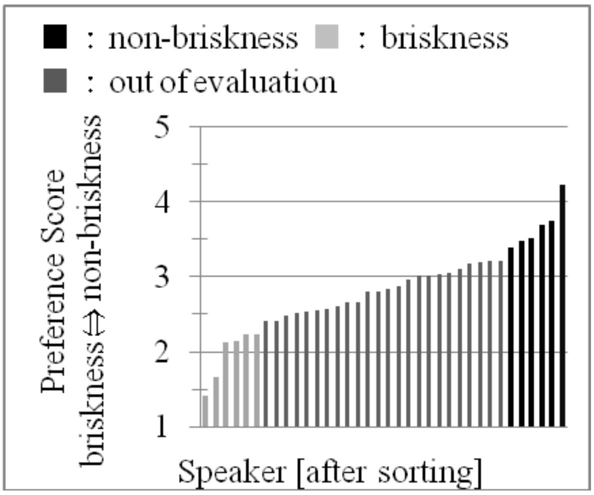


Figure 1: Preference Scores of characteristics of elderly speech

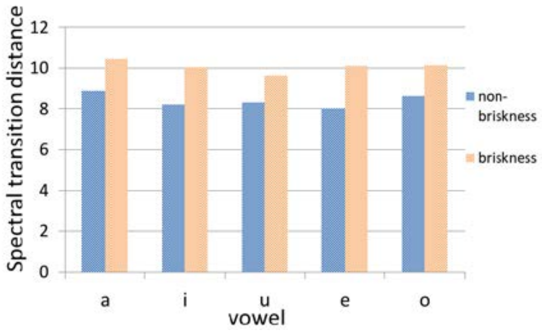
Fig. 2 shows the results of spectral transition distance by degree of briskness. Fig. 2(a) shows the results of initiate phoneme. Fig. 2(b) shows the results of medial phoneme. Fig. 2 shows vowels of non-briskness are lower.

3.2 Difference of spectral transition distance

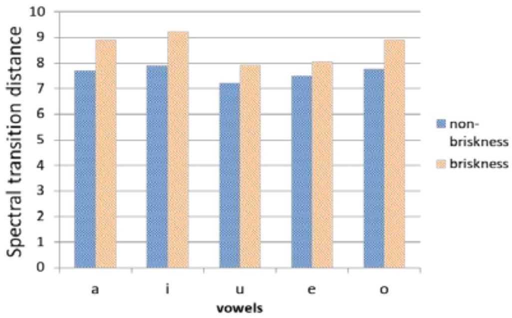
To reveal acoustic feature, we calculated the difference of spectral transition distance. The difference of spectral transition distance is the difference value of spectral transition distance. By calculating it, we reveal vowels of the elderly speech to have non-briskness.

We calculated difference of spectral transition distance by Eq (1). Fig. 3 shows the results of difference of spectral transition distance about the elderly speech by degree of briskness. Fig. 3(a) shows the results of initiate phoneme. Fig. 3(b) shows the results of medial phoneme. Fig. 3(a) shows that /a/, /i/ and /o/ have many differences. Fig. 3(b) shows that vowels expect /u/has many differences.

From this result, /a/, /i/ and /o/ are difficult to utter for the elderly people in beginning to talk. And in midstream to talk, the pronunciation of the elderly people is not good because the movement of the mouth and tongue becomes slow.

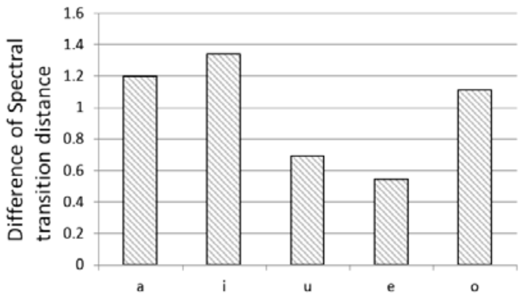


(a) Initiate phoneme

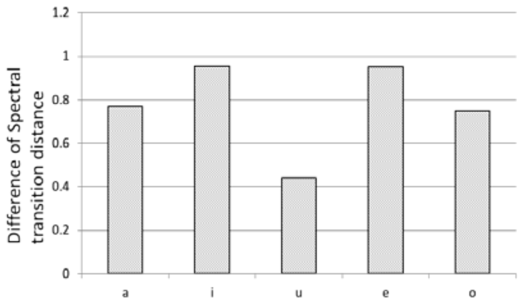


(b) Medial phoneme

Figure 2: Spectral transition distance by degree of briskness



(a) Initiate phoneme



(b) Medial phoneme

Figure 3: Difference of Spectral transition distance about non-brisk speech and brisk speech

Relationship between the absolute difference of spectral transition distance and position of articulation

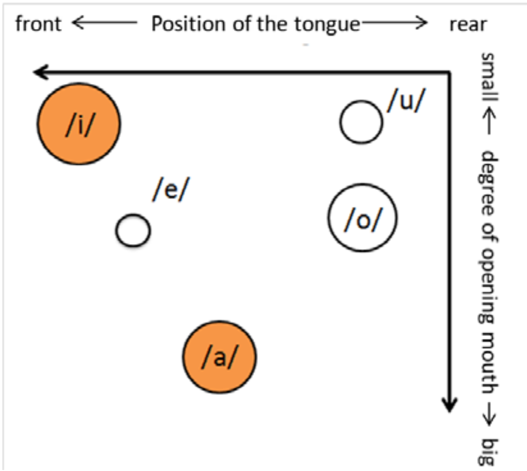
We plotted Fig. 4 about position of articulation and difference of spectral transition distance because of revealing relationship between briskness and an utterance action.

Fig. 4 is the relationship between the absolute difference of spectral transition distance and position of articulation. This radius of the circle is determined as the absolute value of the difference of transition distance. Fig. 4(a) shows the results of initiate phoneme. Fig. 4(b) shows the results of medial phoneme. About position of articulation, we referred to previous researches [4].

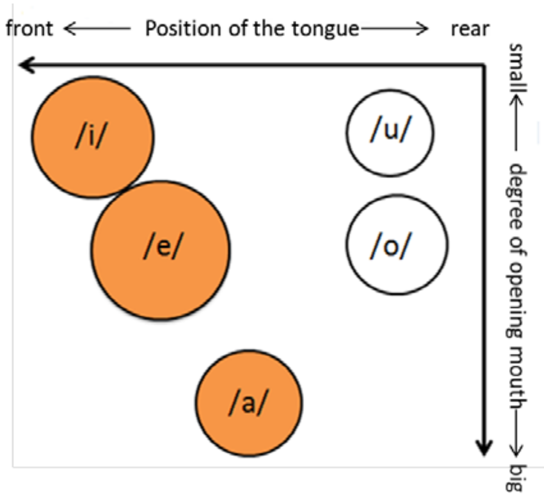
We found that the vowels need large movement of mouth and tongue such as /a/, and /i/ has the big distance of transition distance. Probably, the elderly speakers with non-briskness are not able to move their tongues and mouths well.

4. The Vowels in the F1-F2 Size

We analyzed speech with non-briskness by linear predictive coding (LPC) with focusing F1 and F2, and compared briskness and speech of non-briskness. LPC is a tool used mostly in audio signal processing and speech processing to represent the spectral envelope and digital signal peak of speeches in a compressed form.



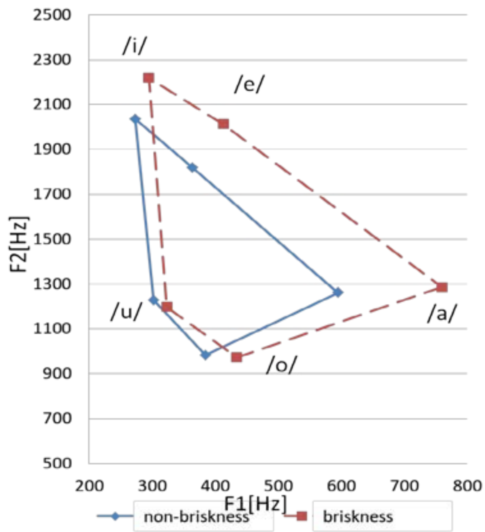
(a) Initiate phoneme



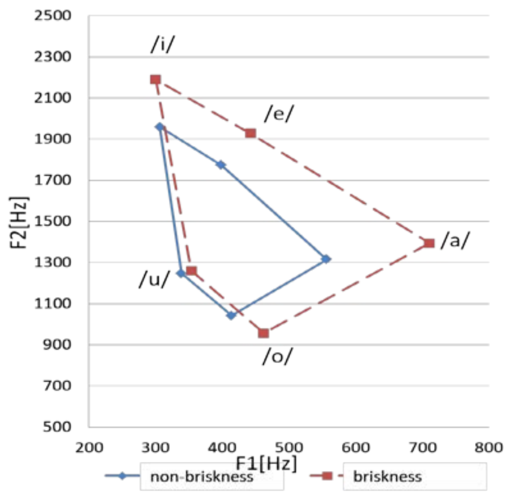
(b) Medial phoneme

Figure 4: Point of articulation and absolute of difference of spectral transition distance

We calculated mean value of F1 and F2 from the 6 speeches with briskness and a speech with non-briskness. Fig. 5 shows the results of calculating F1 and F2. Fig. 5(a) shows the results of initiate phoneme. Fig. 5(b) shows the results of medial phoneme. According to Fig. 5, the vowels expect /u/ with non-briskness converged to /u/. We estimated that due to a decrease in muscle strength with age, the movement of the tongue and mouth become dull, so the vowels become closer to /u/.



(a) Initiate phoneme



(b) Medial phoneme

Figure 5: F1-F2 size of briskness and non-briskness speech

5. Shift of F1 and F2 in the Non-briskness Speech

We shifted speeches with non-briskness to make it close to the briskness speeches about F1 and F2.

5.1 Method of shift

Fig.6 shows the process of the shift of F1 and F2. First, the speeches are divided into a spectrum envelope and a harmonic structure by the LPC. Second, F1 and F2 are calculated from the spectrum envelope. Third, we shift the shift value by F1 and F2. We explain about the shift value Chp5.2. Finally, the shift speeches are created by synthesizing the shifted formant frequency and a harmonic structure. In this case, synthesizer causes distortion to shift speech by error.

5.2 Calculation of the shift value

A shift value was derived by subtracting the mean of the F1 or F2 of the speech with non-briskness from the F1 or F2 on six speeches with briskness. Table.2 shows shift value of one non-brisk speaker. Fig.7 shows that the speeches of before and the after shift. The points of Fig.7 show the formant frequency. The joints of the vowel are revised by manual operation to get smooth. Compared with Fig. 7(a) and Fig. 7(b) , F1 and F2 are shifted above by our method.

6. Verification of Shift Speech by Listening Test

We verified shift speech by listening tests. The method of comparison is based on the paired comparison called Scheffe.

6.1 Listening test

In order to verify the effect of improving intelligibility in this shift method, we arranged “before shift speech", "shift speech" and "non-shift speech”. The “non-shift speech" is the speech which was synthesized without shifting the F1 and F2. And it was used to confirm the influence of the distortion due to synthesis. We performed listening experiments for the speech of three types. The method of comparison is based on the paired comparison called Scheffe. The evaluation words are 10 and contain a few vowels from the database of 543 words (table. 3.). The subjects were 17 adult males.

6.2 Evaluation method

First, subjects listen to two speeches. Two speeches are two of three about processing method at random. Second, subjects evaluate two speeches at five grades about degree of briskness. The evaluation word is intelligibility. Fig. 8 shows about 5 grades of degree of briskness.

6.3 Results of listening test

We calculated the mean of preference degrees and the 99% confidence interval using the method of paired comparison of Scheffe. Fig.9 shows the result of listening experiments. Fig6 (a) shows the comparison between “shift speech” and “non-shift speech”. Fig9 (b) shows comparison between “before shift” and “after shift”. The top side is good for the former. The bottom side is good for the latter.

From Fig. 6(a), the six words, /AOAO/, /MYAKU-MYAKU/,/METYA-METYA/,/PATI-PATI/, /OIOI/ and /UIUISII/, have the meaning of the difference, and "shift speech" is good. These words need large move of tongue and mouth when we try to pronounce. And it might be difficult for the elderly people. Therefore, it was effective for improving the intelligibility by using this formant frequency shift method remarkably. On the other hands, the words, /GYOKUHO/ and /IMEEZI/, do not have significance. It is assumed that these words do not need large move. And it is easy to pronounce for the elderly people. This result is consistent with the result of Chapter 3.

From Fig. 6(b), the five words, /AOAO/, /MYAKU-MYAKU/, /PATI-PATI/ and /OIOI/, have the meaning of the difference. Compared Fig. 6(a) with Fig.6(b), the result of /METYA-METYA/ is different. It is assumed that noise of synthesis affected this result.

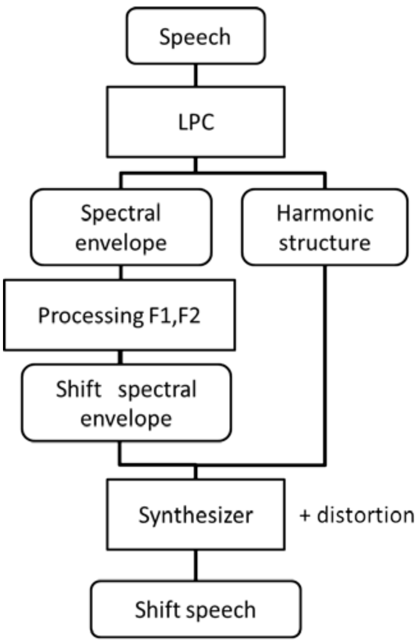
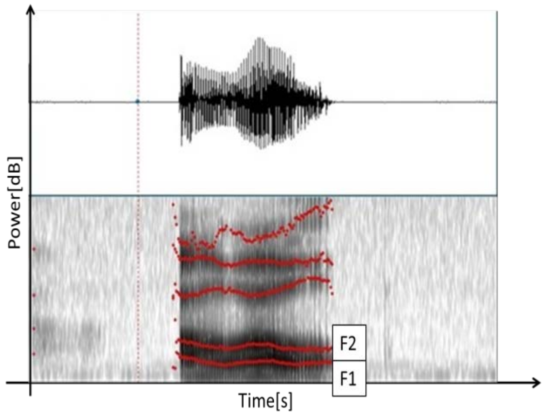
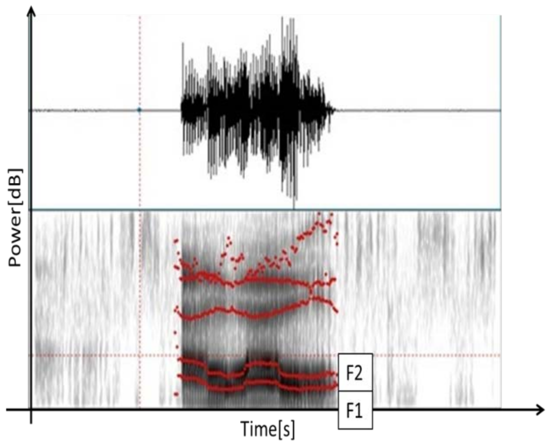


Figure 6: Flowchart of formant frequency shift method

	Medial phoneme		Initial phoneme	
	F1[Hz]	F2[Hz]	F1[Hz]	F2[Hz]
/a/	119	111	100	30
/o/	7	−89	6	−171
/u/	−0.1	144	4	192
/i/	−11	293	18	263
/e/	17	210	43	210



(a) Before processing



(b) After processing

Figure 7: Wave form and Spectrogram of shift non-briskness speech about /AOAO/

Table 3: Evaluation speeches

	/AO AO/	/MYAK U-MYA KU /	/PATI-PATI/	/METYA-METYA/	/GYOK UHO/
	/OI OI/	/PEKO-PEKO/	/UIUI SII/	/HUTEBU TESII/	/IMEE ZI/

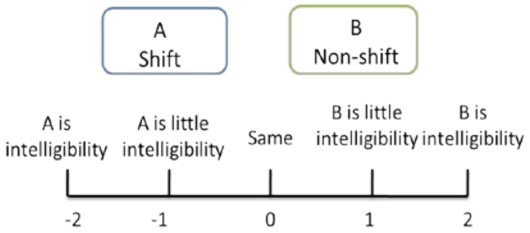
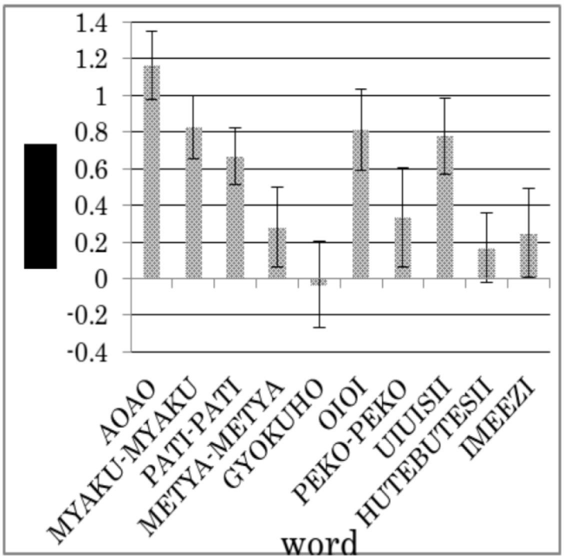


Figure 8: five grades of evaluation about intelligibility



(a)"shift speech"-"non-shift speech"

7. Conclusion

In this study, we analyzed the elderly speeches based on acoustic features, and examined the effect of improving speech with non-briskness using the shifted formant frequency.

Analyzing speech with non-briskness and briskness, it was found that there was a relationship between difference of transition distance and position of articulation. It is difficult for the elderly people to utter /a/ and /i/, because they need large movement of mouth and tongue. In addition, the vowels are converged to /u/ about F1 and F2. The shift method of F1 and F2 for speech with non-briskness improved intelligibility in the words with the large movement of mouth and tongue. In future works, we will investigate analysis and how to improve acoustic features of consonants in the elderly speech.

Acknowledgments

This work was partially supported by Grant-in-Aid for Science Research (No.23560459) from the Japan Society for the Promotion of Science.

References

[1]. T.Miyazaki et al., "Study on auditory impressions of elderly speech," Proc. ASJ meeting, pp. 427-428, 2008.

[2]. Y.Yuda et al.," Investigation on variation of the initial phonemes of Japanese words due to aging", ICISIP2013, pp49-52

[3]. D.Harada et al., "Analysis of ill-articulated speech by elderly speakers based on temporal transition of amplitude spectrum" Proc.NCSP,

[4]. S.Itabashi et al., "Sound engineering(Onsei kougaku)" 2005,pp6-15

Yuto Tanaka He was born in Fkuoka, Japan, on July 6, 1989. He received a M.E. degree in engineers from Kyushu Institute of Technology in 2014. He is currently a doctoral course student at Graduate school of Engineering, Kyushu Institute of Technology. He is a member of IEICE.

Hiroaki Igaue He was born in Ehime, Japan, on November, 1990.During 2009-2013, he was enrolled at the Kyusyu Institute of Technology, Japan. Since 2013, he has been a master degrees student in Kyusyu Institute of Technology, Japan.

Mitsunori Mizumashi He received a Bachelor degree in design from Kyushu Institute of Design and a Ph.D degree in information science from Japan Advanced Institute of Science and Technology (JAIST) in 1995 and 2000, respectively. He is currently an associate professor at Kyushu Institute of technology. His research interests include acoustic information processing and statistical signal processing. He is a member of AES,ASA, ASJ, IEEE, IEICE, and RISP.

Yoshihisa Nakatoh He received his Ph.D degree from shinshu University in 2007. He worked for Sharp Corporation from 1986 to 1989 and worked for Matsushita Electric Industrial Co. Ltd from 1991 to 2008. He is currently a professor at Kyushu Institute of technology. His research interests include speech recognition, audio coding, hearing aid and accessibility technology. He is a member of ASJ, IEICE, IPSJ.