J. Acoust. Soc. Jpn. (E) 18, 6 (1997)

Acoustical and perceptual cues of the palatalized articulation of /s/

Takayuki Arai,* Keiko Okazaki,** Setsuko Imatomi,*** and Yuichi Yoshida****

*International Computer Science Institute, 1947 Center Street, Berkeley, CA 94704, USA
**Department of Plastic and Reconstructive Surgery, School of Medicine, Showa University, 1-5-8, Hatanodai, Shinagawa-ku, Tokyo, 142 Japan
***Department of Speech Therapy, Fujigaoka Rehabilitation Hospital, Showa University, 2-2-1, Fujigaoka, Aoba-ku, Yokohama, 227 Japan
****Department of Electrical and Electronics Engineering, Sophia University, 7-1, Kioi-cho, Chiyoda-ku, Tokyo, 102 Japan

(Received 9 July 1996)

Palatalized articulation (PA) is frequently observed in speech uttered by postoperative cleft palate patients. Provided the acoustical and perceptual cues of PA can be found, speech therapists will be able to use these cues to diagnose PA non-invasively and objectively. We tested human perception of certain synthetic sounds to verify the cues of the PA of /s/ in Japanese. To synthesize the fricatives, we modified the center frequency and the bandwidth of a complex-conjugate pole pair of an all-pole filter obtained from the linear predictive analysis of the PA of /s/. First, we shifted the center frequency from 1,000 to 3,000 Hz, while the relative bandwidth, or Q factor, was fixed at 10. Subsequently, we shifted the Q factor from 1 to 10, while the center frequency was fixed at 1,800 Hz. The results of a perceptual experiment involving nine speech therapists were conclusive that fricatives having a peak between 1,600 and 2,400 Hz tend to be identified as the PA of /s/, and fricatives having a peak at 1,800 Hz with the Q factor >5, tend to be identified as the PA of /s/. The two-tube model also showed that a peak around 2 kHz characterizes the PA of /s/.

Keywords: Cleft palate speech, Palatalized articulation, Acoustic analysis, Synthetic speech, Linear predictive coding

PACS number: 43. 70. Dn

1. INTRODUCTION

After cleft palate surgery many cleft palate patients obtain good velopharyngeal function; however, some misarticulations may remain, for example, palatalized articulation (PA), nasopharyngeal articulation and lateral articulation.[1-4]

PA is an abnormal articulation seen in dental and alveolar sounds, such as /t/ or /s/ in Japanese.[2] Japanese /t/ and /s/ are normally produced by the tongue-tip and teeth (/t/) or the tongue-tip and alveolar ridge (/s/). Okazaki reported that these sounds uttered by cleft palate patients are produced by the elevated tongue dorsum and palate. They observed the articulatory tongue movements of PA using a dynamic palatograph and cineradiograph.[2] Okazaki also found tongue-palate contact in the posterior portion of the palate.[5]

Speech therapists usually diagnose PA by auditory analysis and visual observations of the tongue movement in clinical examinations. Other possible methods of diagnosis include observing tongue contact with a dynamic palatograph and acoustic analysis by sound spectrographs or computers. These methods will be discussed individually in the following paragraphs.

Auditory analysis is a fairly simple process which can be done during a clinical examination. However, auditory analysis can be unreliable because of the subjectiveness of the judgments. One can compensate for the subjectiveness by having many people conduct the diagnosis.

PA is difficult to diagnose by observation of the tongue movement during clinical examinations because the movement of the tongue is not always visible through the mouth opening.

Another method of diagnosing PA is observing palatal tongue contact with a dynamic palatograph. Observation of the tongue in this way is objective and well suited for identifying cases of the PA of /s/. However, one must make an artificial palate for each patient. This is time consuming and costly.

Another alternative is acoustic analysis using sound spectrographs. Acoustic analyzers using computers are also becoming available today. Acoustic analysis is objective and non-invasive. In order to make this method useful, it is imperative that we find acoustical and perceptual cues that characterize PA. Provided these cues can be found, speech therapists will be able to use the cues to diagnose PA non-invasively and objectively.

In the literature, there are some acoustic analyses for the vowels of cleft palate speech using formant analysis.[6,7]There are fewer analyses of consonants, and a subset of those target the PA of /s/.[8-11]

Wakumoto[9] has reported that the PA of /s/ and the normal articulation of /s/ was distinguishable by the differences between high sub-band level (5.0-7.5 kHz) and low sub-band level (1.5-4.0 kHz) on the consonant spectrum envelope.

Okazaki *et al.*[8] have reported on spectrum characteristics of the PA of /s/ using a sound spectrograph for the acoustic analysis. They found that the highest peak of energy for /s/ is at 2-3 kHz in most cases of the PA group, but over 5 kHz in the normal articulation group. They also report that the range of frequencies at which substantial acoustic powers for /s/ is below 4 kHz in the PA group, but over 4 kHz in the normal articulation group. The PA of /s/ in Okazaki's experiment also shows different power spectra from /ʃ/ in the normal articulation group.

In our previous study, we determined to specify perceptual cues of the PA of /s/ by sounds synthesized by the all-pole filter with a single complex-conjugate pole pair.[12] We shifted the center frequency of the pole pair to see how the location of the spectral peak affects perception of the sounds by speech therapists. The center frequency was shifted at intervals of 500 Hz from 2,000 to 6,000 Hz, while the bandwidth was fixed at 100 Hz. The results of the study show that fricatives having a sharp spectral peak at 2,000 Hz tend to be identified as the PA of /s/.

The results of this study brought forth new questions: What will happen when the center frequency is shifted more finely? When the center frequency is shifted from lower frequency? When the bandwidth is shifted?

To answer these questions, further experiments and analyses were conducted. In the present study, we analyzed and synthesized the PA of /s/ in Japanese with linear predictive coding (LPC). To synthesize the fricatives, we modified the center frequency and the relative bandwidth of a local peak around 2 kHz. Then perceptual experiments were conducted using the synthesized sounds. We also investigated the relation between the zero-pole location and the vocal-tract configuration using a highly idealized two-tube model.

2. METHOD

2.1 Analysis by LPC

We began with speech identified as the PA of /s/ which was uttered by three different postoperative cleft palate children. The children A, B and C are 5, 5, and 10 years old, respectively. The original speech sounds were recorded on an analog tape. We digitized the sounds with 16 kHz sampling rate and 16 bit quantization.

We then performed an eighth-order LPC analysis. This order was chosen to capture major spectral components of the PA of /s/.[12] Figure 1 shows the LPC spectra for each child's speech. The LPC coefficients were calculated by taking the Hamming window with a length of 16 ms.

2.2 Synthesis by LPC

An eighth-order all-pole model was used to synthesize the fricatives for the experiments. The pole locations in the z-plane are shown in Fig. 2. The

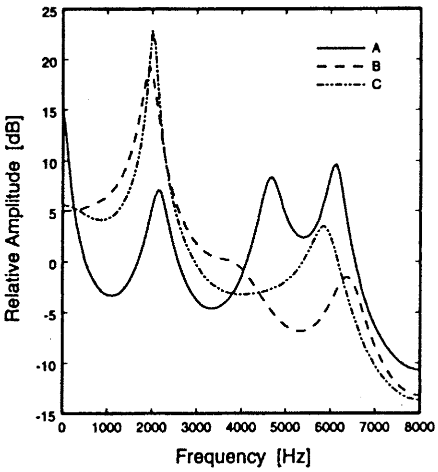

Fig. 1 Spectra of the palatalized articulation of /s/ uttered by three children A, B and C.

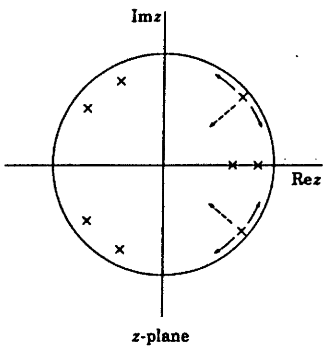

Fig. 2 Pole locations of the filter for the synthesis of fricatives. The poles at 1,800 Hz were shifted along the arrows.

frequency and the bandwidth of each pole are listed in Table 1. These poles were originally calculated from a typical PA of /s/ by LPC analysis. The real poles performed to adjust the tilt of the spectrum. The center frequency, F_P, and the bandwidth, B_P, of the complex-conjugate pole pair P were variables.

The synthesized fricatives were followed by the vowel /a/. The vowel was synthesized by a fourth order all-pole model from the first and second formants shown in Table 2.

To implement the filter, PARCOR coefficients were used. The filter was excited by white noise to synthesize the fricative sounds and by a sequence of impulses to synthesize the vowel sounds. The duration of the fricatives and vowels were 120 ms and 240 ms, respectively.

Table 1 Poles for the synthesis of fricatives.

Pole	Frequency [Hz]	Bandwidth [Hz]
Real pole 1	0	400
Real pole 2	0	1,400
Complex-conjugate pole pair 1	5,200	800
Complex-conjugate pole pair 2	6,400	800
Complex-conjugate pole pair P	F_P	B_P

Table 2 The first and second formants for the synthesis of /a/.

	Center frequency [Hz]	Bandwidth [Hz]
F_1	1,100	150
F_2	1,800	200

Table 3 Conditions for the synthesis.

Sampling	16 kHz
Quantization	16 bit
Duration of the fricative	120 ms
Duration of the vowel	240 ms
Fundamental frequency of the vowel	320 Hz

Because our synthesis was intended to mimic children's speech, we used a fundamental frequency of 320 Hz for the vowels. We adjusted jitter and shimmer for the sequence of voice source impulses to avoid unnatural quality of the synthetic speech. The standard deviation of the pitch periods and the amplitude of impulses were both 2% of the mean value of each parameter. Formant transition was considered at the first 40 ms of each vowel.

The ratio of the energy of vowels to that of fricatives was adjusted +16 dB. The syllables were synthesized on a workstation with 16 kHz sampling and 16 bit quantization as shown in Table 3. The synthesized sounds were then recorded on audio tapes.

2.3 Stimulus Conditions

The frequency range of interest for the main experiment was determined during the preliminary experiment. In the preliminary experiment we studied three speech therapists' perception of the synthesized syllable /sa/, having shifted only F_P. We shifted F_P at intervals of 100 Hz from 1,000 to 3,000 Hz, while B_P was fixed at 100 Hz. We found that the fricatives having a peak around 1,800 Hz tended to be identified as the PA of /s/.

In the main experiment, the center frequency and the relative bandwidth were shifted independently. Since the human auditory system is often discussed in terms of critical band, it is useful to use the relative bandwidth, or the Q factor[13] instead of B_P. The Q factor of the pole P is defined in the following ratio:

$$Q_P = F_P/B_P$$

First, we shifted F_P at intervals of 100 Hz from 1,000 to 3,000 Hz, while Q_P was fixed at 10. Twenty-one different speech samples were synthesized to investigate the effects of the center frequency.

Second, we shifted Q_P. Since the center frequency of 1,800 Hz was mostly identified as the PA of /s/ in the preliminary experiment, F_P was fixed to 1,800 Hz. We shifted Q_P at intervals of 1 from 1 to 10. Ten speech samples were prepared to investigate the Q-factor effects.

The center frequencies and the bandwidths of the rest of the poles were fixed as in Table 1. A complete set of 30 speech samples (21 emphasizing the center frequency, 10 emphasizing the Q factor, and one sample in common to both groups) was presented to subjects during the sessions.

2.4 Procedure

We used the method of constant stimuli with stimuli presented in random order. Each stimulus was repeated three times.

After hearing a speech sample, the speech therapists were prompted to select an option from Table 4 based on what they heard. The options are "the PA of /s/," "highly close to the PA of /s/," "slightly close to the PA of /s/," or "not the PA of /s/." Each option was associated with a score as listed in Table 4.

Each subject participated in two sessions. The same set of speech samples were used in the second session, but in a different order.

Table 4 The options and their scores for the perceptual experiment.

Options	Scores
the PA of /s/	3
highly close to the PA of /s/	2
slightly close to the PA of /s/	1
not the PA of /s/	0

2.5 Subjects

A total of nine speech therapists participated in the study. The subjects were all female, with a range of 4-29 years experience (the average was 13 years).

2.6 Acoustic-tube Model

To determine whether the abnormal place of articulation generates the cue of the PA of /s/, we modeled the PA of /s/ based on an acoustical theory. According to an acoustical theory of speech production, the spectra of fricatives can be characterized by zeros and poles. The location of those zeros and poles are dependent on the shape of the vocal tract and on the location of the sound source. We estimated the location of zeros and poles using a highly idealized two-tube model of the vocal tract as shown in Fig. 3. This simplified model gives us a reasonable estimation of the zero-pole location for voiceless fricatives.[14]

Figure 4 shows the equivalent circuit of this two-tube model, where P_t is the sound pressure generated

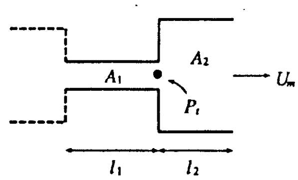

Fig. 3 A two-tube model for the production of the fricatives. We assume that the constriction is narrow enough, so the cavities posterior to the constriction (dashed line) are neglected in the simulation.

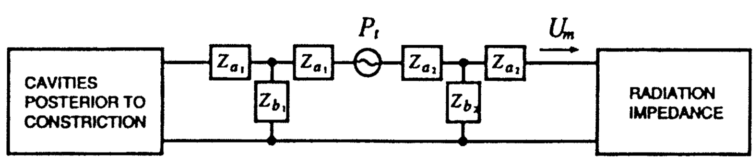

Fig. 4 An equivalent circuit of the two-tube model.

by turbulent flow and U_m is the mouth volume current. In Fig. 4 the first and the second cavities are represented as T-sections, where the first cavity is the constriction and the second cavity is the cavity anterior to the sound source. A_i and l_i are the cross-sectional area and the length of the i-th cavity ($i=1, 2$).

3. EXPERIMENTAL RESULTS

3.1 Effect of the Center Frequency

Figure 5 shows the mean scores for the main experiment involving the nine speech therapists' perception of the synthesized syllable /sa/. The horizontal axis in this figure is F_P, which we shifted at intervals of 100 Hz from 1,000 to 3,000 Hz. The doted line at 1.5 is the mid value within the range of scores.

In Fig. 5 we divided the results into two groups, A and $\bar{A}$. Group A is the subset of the results having mean scores of 1.5 or greater, and Group $\bar{A}$ contains mean scores less than 1.5. The Wilcoxon test revealed that the mean score of Group A was significantly larger than that of Group $\bar{A}$ at the 5% level. This shows that fricatives having a peak in the range of 1,600–2,400 Hz tend to be identified as the PA of /s/.

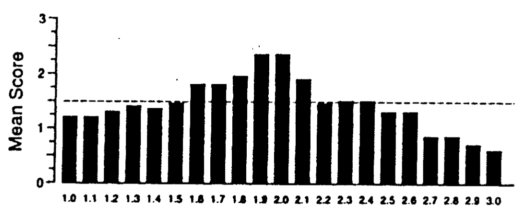

Fig. 5 Results of the perceptual experiment for the center frequency.

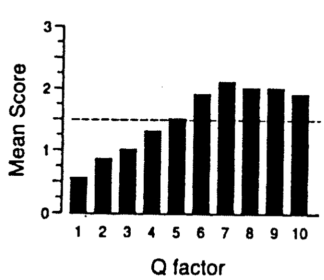

Fig. 6 Results of the perceptual experiment for the Q factor.

3.2 Effect of the Q Factor

Figure 6 shows the averaged scores when Q_P was shifted from 1 to 10. The doted line indicates a score of 1.5. As shown in Fig. 6, the mean score was greater than 1.5 when the fricatives have a peak at 1,800 Hz with $Q_P>5$.

We divided the results in Fig. 6 into two groups, B and $\bar{B}$. Group B is the subset of the results having mean scores greater than 1.5; Group $\bar{B}$ contains mean scores of 1.5 or less. The Wilcoxon test revealed that the mean score of Group B was significantly larger than that of Group $\bar{B}$ at the 1% level. This shows that fricatives having a peak at 1,800 Hz with $Q_P>5$ tend to be identified as the PA of /s/.

3.3 Estimation of the Zero-pole Location

We estimated the zero-pole location with an acoustic-tube model. If the constriction is narrow enough, the output spectrum is virtually unchanged when the cavities posterior to the constriction are neglected.[14] If the radiation impedance is assumed to be zero, the driving point admittance[15] is

$$\frac{U_m}{P_t}=\frac{Z_{a2}//Z_{b2}}{Z_{a2}(Z_{a1}+Z_{a2}+Z_{a1}//Z_{b1}+Z_{a2}//Z_{b2})} \quad (1)$$

where // indicates the parallel connection of two impedances. And therefore,

$$\frac{U_m}{P_t}=\frac{k}{\sinh \gamma_2 l_2\left(\tanh \gamma_1 l_1+\frac{A_1}{A_2}\tanh \gamma_2 l_2\right)} \quad (2)$$

where k is a constant and γ_i is the propagation constant[16] at the i-th cavity.

The place of articulation of the PA of /s/ is further back compared to the normal articulation of /s/.[2] In the case of (A) in Table 5, the constriction in Fig. 3 is located just behind the mouth opening, and it approximates the normal articulation of /s/. These dimensions are based on Heinz (1961).[14] In the case of (B) in Table 5, the constriction is located further back and approximates the PA of /s/. These dimensions are based on Okazaki (1980)[2] and Heinz (1961).[14]

Table 5 Dimensions of the cavities for the two-tube model.

	Area [cm²]		Length [cm]	
	A_1	A_2	l_1	l_2
(A)	0.2	6.0	2.8	0.8
(B)	0.2	8.0	2.8	4.3

Table 6 Location of zeros and poles of the two-tube model.

	Zeros [Hz]	Poles [Hz]			
(A)	3,040	0			5,990
(B)	3,040	0	2,000	5,750	6,250

When we use the dimensions of the cavities in Table 5, the location of zeros and poles up to 8 kHz are listed in Table 6. The values in Table 6 are calculated under the assumption that the acoustic tube is lossless and the velocity of the sound is 340 m/s.

4. DISCUSSION

The spectra of the PA of /s/ in Fig. 1 suggest that a local peak around 2 kHz is a crucial cue. They are consistent with the previous study.[8] The results of the perceptual experiment (Figs. 5 and 6) show that the sharp peak ($Q_P>5$) around 2 kHz characterizes the PA of /s/. The results confirm our previous claims[12] but also give us more information.

In both results of our previous and present studies, the highest score is located at 2 kHz. The present result, however, shows that the scores are gradually decreasing towards lower or higher frequency.

According to a measurement of the critical bandwidth of human auditory system,[13] the Q factor in the frequency range of 1-3 kHz is approximately 6. When we shifted F_P at intervals of 100 Hz from 1,000 to 3,000 Hz, Q_P was fixed at 10. Therefore, B_P was always less than the critical band. Since the mean score in Fig. 6 is saturated when $Q_P>5$, we can say that fricatives tend to be identified as the PA of /s/ when the peak is within the critical band.

In both cases (A) and (B) in Table 5 and 6 there is a common pole at 0 Hz and a common zero around 3 kHz. The pole at 0 Hz contributes to adjust the tilt of the spectrum. The zero around 3 kHz is located far enough from any of the poles, so that the poles are not affected by the zero. Therefore, an all-pole model may yield a reasonable filter to synthesize the PA of /s/. This zero around 3 kHz can be seen as a dip in spectra of the PA of /s/, and it will be seen as a gap on the spectrogram between 2 and 5 kHz.

There is a single pole in (A) around 6 kHz, while there are two poles around 6 kHz in (B). Those poles in (B) are close to each other, and in both cases (A) and (B), the poles around 6 kHz are similarly emphasizing a high frequency region.

The crucial difference between (A) and (B) is the existence of a pole around 2 kHz. This pole exists in (B) but not in (A), and it seems like this pole characterizes the PA of /s/.

The location of the peak around 2 kHz is sensitive to the parameter l_2. If the place of constriction moves towards the lips and l_2 becomes 1 cm shorter, the center frequency of the peak moves to about 2,600 Hz. On the other hand, if the place of constriction moves towards the glottis and l_2 becomes 1 cm longer, the center frequency becomes about 1,600 Hz. This variation suggests that the range of the center frequency could be more than several hundreds hertz.

This simulation using the two-tube model confirmed that there is no pole around 2 kHz when /s/ is articulated normally, and there is an additional peak around 2 kHz when the place of constriction moves toward the back of the oral cavity. The result of this simulation supports our results of the perceptual experiments.

With the discovery of this cue for PA, speech therapists will be able to diagnose PA more objectively, provided they have the proper tools to analysis speech acoustically.

The identification of this cue will also help speech therapists to chart the progress of their patients. When patients attend training sessions to help them articulate normally, the speech therapists will be able to measure the degree of palatalization across speech training sessions objectively by measuring the Q factor of the peak around 2 kHz.

5. CONCLUSION

To verify the characteristics of the PA of /s/ in Japanese, we analyzed the sounds with LPC analysis and tested human perception of certain syllables synthesized by an all-pole model. The results of the perceptual experiment showed that fricatives having a peak in the range of 1,600–2,400 Hz tend to be identified as the PA of /s/, and that fricatives having a peak at 1,800 Hz with the Q factor >5 tend to be identified as the PA of /s/.

The simulation using a highly idealized acoustic tube model showed that there are no poles around 2 kHz when /s/ is articulated normally, whereas an additional pole appears around 2 kHz when the place of the constriction moves back in the oral cavity. There is also a zero between 2 and 5 kHz, and it can be seen as a dip in the spectrum.

The identification of this cue will aid in the objective diagnosis of the PA of /s/.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of the speech therapists who participated in the perceptual experiment in this study. We also acknowledge advice from Dr. Sharon Oviatt and help by Terri Lander at Oregon Graduate Institute of Science & Technology.

REFERENCES

1) M. E. Morlay, *Cleft Palate and Speech*, 7th ed. (E.& S. Livingston, Edinburgh and London, 1970), pp. 173-210.
2) K. Okazaki, T. Onizuka, M. Abe, and M. Sawashima, "Palatalized Articulation as a Type of Cleft Palate Speech: Observation by Dynamic Palatograph and Cineradiograph," Jpn. J. Logop. Phoniatr. **21**(2), 109-120 (1980) (in Japanese).
3) M. Abe, "Nasopharyngeal articulation: A particular type of articulation error," Jpn. J. Logop. Phoniatr. **28**(3), 239-256 (1987) (in Japanese).
4) M. Kato, "Lateral articulation: Observation of articulatory movement using electro tooth-palatography," Jpn. J. Logop. Phoniatr. **32**(1), 18-22 (1991) (in Japanese).
5) K. Okazaki, M. Kato, and T. Onizuka, "Palate morphology in children with cleft palate with palatalized articulation," Ann. Plast. Surg. **26**, 156-162 (1991) (in Japanese).
6) D. R. Dickson, "An acoustic study of nasality," J. Speech Hear. Res. **5**, 103-111 (1962).
7) M. F. Schwartz, "Acoustic measures of nasalization and nasality," in *Communicative Disorders Related Cleft Lip and Palate*, K. R. Bzoch, Ed. (Little Brown and Company, Boston, 1979), pp. 194-200.
8) K. Okazaki and M. Kato, "Acoustic analysis of palatalized articulation of [s] sounds," Jpn. J. Logop. Phoniatr. **29**(3), 225-231 (1988) (in Japanese).
9) M. Wakumoto, "Quantitative technique for evaluation of Japanese misarticulations—Acoustical characteristics of Japanese palatalized articulation—," Jpn. J. Cleft Palate Assoc. **14**, 21-43 (1989) (in Japanese).
10) T. Fukuda, "Acoustic characteristics of palatalized articulation in cleft palate children and changes following speech therapy," J. Osaka Univ. Dent. Soc. **38**, 8-35 (1993) (in Japanese).
11) Y. Ozawa and K. Okazaki, "Articulatory tongue movements and acoustic characteristics of palatalized articulation in cleft palate patients," Jpn. J. Logop. Phoniatr. **35**, 322-330 (1994) (in Japanese).
12) T. Arai, K. Okazaki, and S. Imatomi, "Palatalized articulation of [s] sounds using synthetic speech," Jpn. J. Logop. Phoniatr. **36**, 350-354 (1995) (in Japanese).
13) B. C. Moore, *An Introduction to the Psychology of Hearing*, 3rd ed. (Academic Press, London, 1989).
14) J. M. Heinz and K. Stevens, "On the properties of voiceless fricative consonants," J. Acoust. Soc. Am. **33**, 589-596 (1961).
15) J. L. Flanagan, *Speech Analysis; Synthesis and Perception* (Springer-Verlag, Berlin, 1972).
16) G. Fant, *Acoustic Theory of Speech Production* (Mouton & Co., The Hague, 1960).

Takayuki Arai received the B.E., M.E., and Ph.D degrees in electrical engineering, from Sophia University, Tokyo, Japan, in 1989, 1991, and 1994, respectively. In 1992-1993 and 1995-1996, he was with Oregon Graduate Institute of Science and Technology (Portland, OR, USA). He is currently a research fellow at the International Computer Science Institute and University of California, Berkeley (Berkeley, CA, USA). His research interests include speech signal processing, speech/language recognition, and acoustic analysis of speech disorders.

Keiko Okazaki received the B.A. degree in psychology from Tokyo Women's Christian College in 1954. In 1974, she joined the Department of Plastic & Reconstructive Surgery, School of Medicine, Showa University. She received Dr. Medical degree from Showa University in 1982. She is interesting in cleft palate speech, especially analysis of articulation disorders.

Setsuko Imatomi received B.A. and M.A. degrees respectively, from Gakushuin University and Sophia University, Japan. In 1991, she was with Chiba Children's Hospital, and in 1992 with Fujigaoka Rehabilitation Hospital, Showa University, as speech therapist. She has joined Sophia University Research Center for Communication Disorders as part-time research assistant, since 1996. Since 1997 she has been with the Department of Plastic and Reconstructive Surgery as speech therapist and research assistant. Her current research interests are in the area of acoustic and phonetic analysis of cleft palate and dysarthric speech.

Yuichi Yoshida was born in Gunma, Japan, on May 2, 1930. He received the B.E. and the Ph.D degrees, from the Univeristy of Tokyo in 1953 and 1968 respectively. He worked at Yaou Denki Co. (Fujitsu General Co., at present) from 1953 to 1960. From 1960 to 1962, he joined the department of electronic engineering, the University of Tokyo, as a researcher. In 1962, he joined the Faculty of Science and Technology, Sophia University, Tokyo, where he is a Professor since 1971. He has been engaged in developments and researches on electronic circuits, communication systems, signal processing and speech recognition.