4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

ANALYSIS FOR PALATALIZED ARTICULATION OF [S] SOUNDS USING SYNTHETIC SPEECH

Takayuki Arai [1] *Keiko Okazaki* [2] *Setsuko Imatomi* [3] *Yuichi Yoshida* [4]

[1,4]Dept. of Electrical and Electronic Engineering, Sophia Univ., Tokyo, JAPAN
[2]Dept. of Plastic and Reconstructive Surgery, School of Medicine, Showa Univ., Tokyo, JAPAN
[3]Dept. of Speech Therapy, Fujigaoka Rehabilitation Hospital, Showa Univ., Yokohama, JAPAN

10.21437/Eurospeech.1995-439

ABSTRACT

Palatalized articulation (PA) is frequently observed in speech uttered by postoperative cleft palate patients. We analyzed the PA of [s] sounds and tested human perception of certain synthetic sounds to verify the characteristics of the PA of [s] sounds in Japanese. After analyzing the PA of [s] with linear predictive (LP) analysis, the mono-syllable /sa/, /su/ and /se/ were synthesized by an all-pole model. To synthesize the fricatives, we shifted the frequency of a complex-conjugate pole pair of a filter from 1000 to 3400 Hz. A perceptual experiment involving three speech therapists was carried out to analyze perception of the three syllables. From the results we concluded that fricatives having a peak around 1800 Hz tend to be identified as the PA of [s].

1. INTRODUCTION

After cleft palate surgery many cleft palate patients obtain normal velopharyngeal function; however some misarticulations may remain [1], such as palatalized articulation (PA) [2], nasopharyngeal articulation [3] and lateral articulation [4]. PA is an abnormal articulation seen in dental and alveolar sounds, typically in [s] or [t] in Japanese [5]. These sounds, normally produced by the tongue-tip and teeth or alveolar ridge, are produced instead by the elevated tongue dorsum and palate, with tongue-palate contact also seen in the posterior portion of the palate [2]. Articulatory tongue movements of the PA were observed using dynamic palatograph and cineradiograph [5].

At present, speech therapists diagnose PA by auditory analysis and visual observations from the mouth opening in clinical examination. If we can find an acoustic cue signaling PA, it will help speech therapists to diagnose PA non-invasively and objectively. 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]. Okazaki [8] has reported on spectrum characteristics of the PA of [s], using a sound spectrograph for the acoustic analysis: 1) the highest peak of energy for [s] sounds were at 2 – 3 kHz in most cases of the PA group, and over 5 kHz in the normal articulation group; 2) the range of frequencies at which substantial acoustic powers were seen for [s] sounds falls below 4 kHz in the

[1]Currently at Dept. of Computer Science and Engineering, Oregon Graduate Institute, Portland, Oregon, U.S.A. (E-mail: arai@cse.ogi.edu)

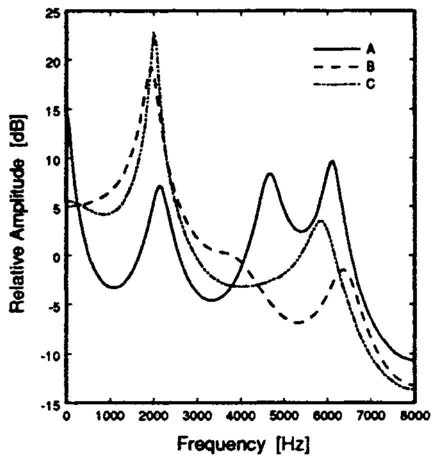

Figure 1. Spectra of the palatalized articulation of [s] sounds uttered by three children A, B and C.

PA group, but was over 4 kHz in the normal articulation group; 3) the PA of [s] sounds also showed different power spectra from [ʃ] sounds in the normal articulation group. Wakumoto [9] has reported that the PA of [s] and the normal articulation of [s] was distinguishable by the differences between high (5.0 – 7.5 kHz) and low (1.5 – 4.0 kHz) sub-band level on the consonant spectrum envelope.

In this study, we analyzed the PA of [s] sounds with linear predictive (LP) analysis and tested human perception of certain sounds synthesized by an all-pole model to verify those characteristics of the PA of [s] sounds in Japanese.

2. PRELIMINARY EXPERIMENT

In the mono-syllable /su/ uttered with normal articulation, part of the fricative sound [s] was replaced by synthetic noise. The following three models were used to synthesize [s]. The first model was a bandpass filter which had a pass band in a specific frequency range. The cutoff frequencies were shifted in intervals from low to high, while the bandwidth was fixed. The second model was an all-pole model with second-order LP analysis. To implement this filter partial correlation (PARCOR) coefficients were used. The frequency of the pole pair was shifted at intervals of 500 Hz from 2 to 6 kHz, while its bandwidth was fixed at 100 Hz. The third model was an all-pole model using higher order LP analysis for a typical PA of [s]. Each filter was excited by white noise to synthesize the fricatives.

Table 1. Poles for synthesis of fricatives

Pole	Frequency [Hz]	Bandwidth [Hz]
Real Pole 1	0	400
Real Pole 2	0	1200
Complex Conjugate Pole Pair 1	5200	800
Complex Conjugate Pole Pair 2	6400	800
Complex-Conjugate Pole Pair P	p	100

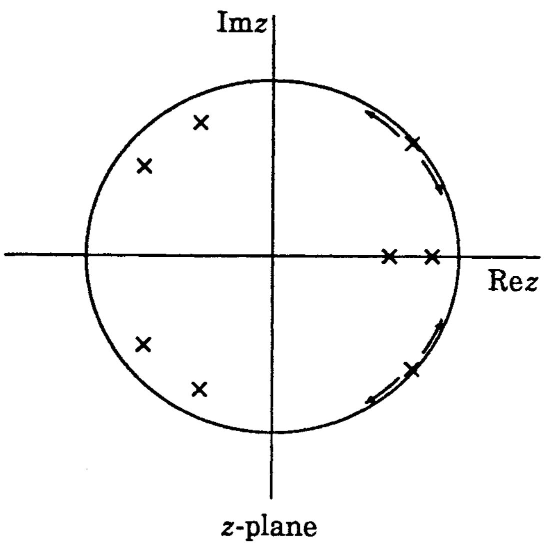

Figure 2. Pole locations of the filter for the synthesis, where p = 1800 Hz.

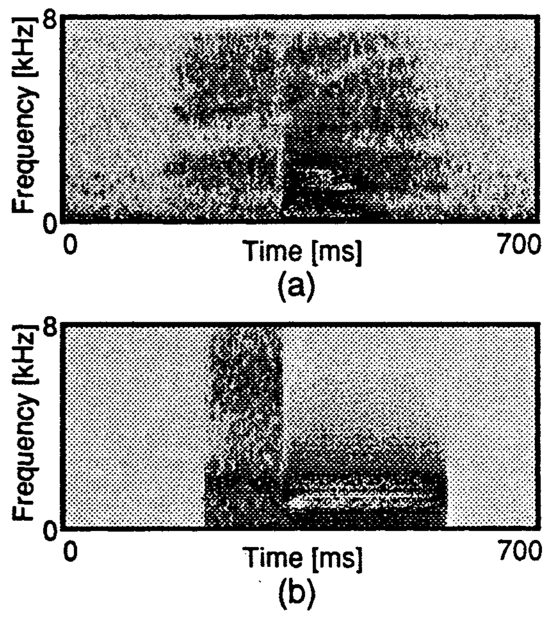

Figure 3. The sound spectrograms of /sa/ for (a) speech uttered by the child A and (b) synthesized speech with p = 1800Hz.

The hearing discrimination of nine speech therapists formed the data for the perceptual experiment; they were each requested to indicate what they heard. Their replies were categorized as: "[s]," "[ʃ]," "the PA of [s]," or "other." From the results we concluded: 1) the first model was not appropriate for synthesizing the PA of [s]; 2) fricatives, which had a peak less than 3 kHz, tended to be identified as the PA of [s] when synthesized by the second model; and 3) fricatives synthesized by the third model tended to be identified as the PA of [s] when the bandwidth of the peak near 2 kHz was less than about 500 Hz.

3. ANALYSIS OF THE PA OF [S]

The speech uttered by three different postoperative cleft palate children and identified as the PA of [s] were analyzed with LP analysis. The spectra of children A, B and C from LP coefficients were shown in Fig. 1. An additional peak around 2 kHz seems to characterize the PA of [s]. This peak was not observed in the spectra of the normal articulation of [s]. The frequencies of the peaks are located around 2 kHz, which is consistent with results obtained in a previous study [8].

4. SYNTHESIS OF FRICATIVES

An eighth-order all-pole model was used to synthesize fricatives in the main experiment. The pole locations in the z-plane are shown in Fig. 2; the frequency and the bandwidth of each pole are listed in Table 1. These poles are originally calculated from a typical PA of [s] sound by LP analysis. The real poles performed to adjust the tilt of the spectrum. The frequency p of the complex-conjugate pole pair P was variable.

The synthesized fricatives are followed by vowels /a/, /u/ or /e/ synthesized by a fourth order all-pole model from the first and the 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. Because we are assuming speech by children, we used a fundamental frequency of 320 Hz for the vowels; the duration of the fricatives and the vowels were 120 ms and 240 ms, respectively. We considered jitter and shimmer for the sequence of impulses, that is, standard deviation of the pitch periods and the amplitude of impulses are both 2% of the mean value of each parameter. Formant transient was considered on the first one sixth of each vowel (40 ms). The ratio of the energy of vowels to that of fricatives was adjusted +16 dB. Those fricatives were synthesized on the work station under conditions with 16 kHz sampling and 16 bit quantization as shown in Table 3. Finally, those sounds were recorded on the audio tapes. The sound spectrograms of /sa/ for speech uttered by the child A in the Section 3 and synthesized speech with p = 1800 Hz are shown in Fig. 3.

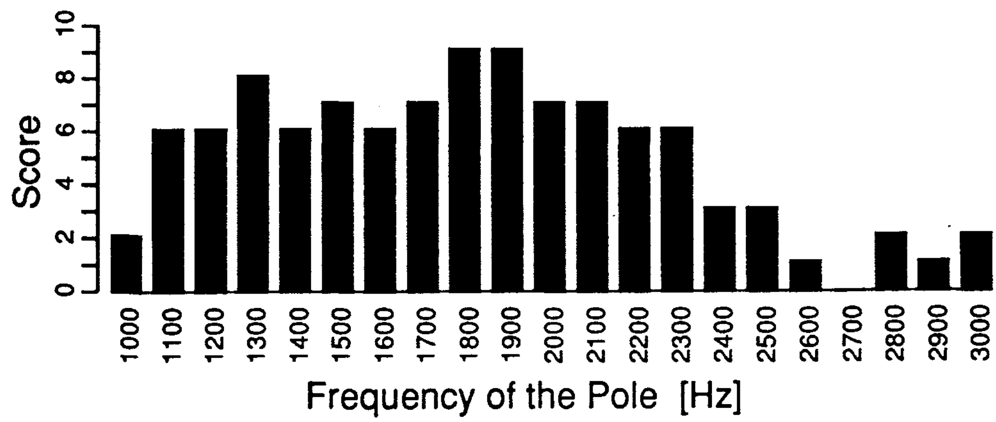

Figure 5. Results of Exp. 2

Table 2. The first and the second formants for /a/, /u/ and /e/.

Vowels			[Hz]
/a/	F1	Center Frequency	1100
		Bandwidth	150
	F2	Center Frequency	1800
		Bandwidth	200
/u/	F1	Center Frequency	500
		Bandwidth	50
	F2	Center Frequency	1600
		Bandwidth	50
/e/	F1	Center Frequency	700
		Bandwidth	50
	F2	Center Frequency	2800
		Bandwidth	50

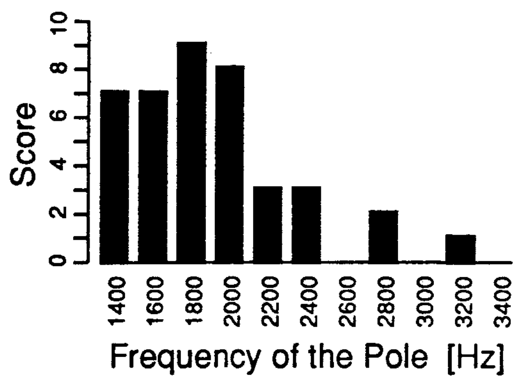

Figure 4. Results of Exp. 1

5. PERCEPTUAL EXPERIMENTS

5.1. Experiment 1

The perceptual experiment of a speech therapist was carried out for the syllables /sa/, /su/ and /se/. The frequency of the pole pair P was shifted at intervals of 200 Hz from 1400 to 3400 Hz, while the rest of the poles were fixed. Thirty-six samples, including samples which do not have the pole pair P, were randomly played back three times. The same experiment was done twice. After hearing a speech sample, the speech therapist was prompted to select an option from Table 4 based upon what he/she heard. The samples, which do not have the pole pair P, have been selected as the normal articulation of [s] sounds. The [ʃ] was added to the options, but the [ʃ] was never selected as an answer. Figure 4 shows the results of the test for each p. The total scores in the figure were calculated as summation of a score shown in Table 4 for each selected option. The fricatives, which have a peak less than 2400 Hz, tend to be identified as the PA of [s]. There was no significant difference observed in the results of the three different syllables, /sa/, /su/ and /se/.

5.2. Experiment 2

In order to investigate effects caused by a slightly lower frequency range, a second perceptual experiment was conducted. In this experiment there were three speech therapists and the only syllable studied was /sa/. The frequency of the pole pair P was shifted at intervals of 100 Hz from 1000 to 3000 Hz, while the rest of the poles were fixed. After hearing each speech sample consisted of an utterance of syllable containing a fricative, the speech therapists were prompted to select an option from Table 5 based upon what they heard. Twenty-one samples were randomly played back three times. The same experiment was done twice. Figure 5 shows the results of the test for each p. The total scores in the figure were calculated as summation of a score shown in Table 5 for each selected option. The fricatives, which have a peak in the range of 1400 – 2400 Hz, tend to be identified as the PA of [s].

Table 3. Conditions for Exp. 1 and Exp. 2

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

Table 4. The options and their scores for Exp. 1

Options	Scores
the normal articulation of [s]	0
close to the normal articulation of [s]	0
the normal articulation of [ʃ]	0
close to the normal articulation of [ʃ]	0
the PA of [s]	2
close to the PA of [s]	1
other	0

Table 5. The options and their scores for Exp. 2

Options	Scores
the PA of [s]	2
close to the PA of [s]	1
other	0

6. DISCUSSION

To verify the characteristics of the PA of [s] sounds in Japanese, we analyzed the sounds with LP analysis and tested human perception of certain sounds synthesized by an all-pole model. In fact, we needed zeros to synthesize fricative sounds, because the noise source of the sounds is located at the anterior portion of the oral cavity [12]. Our results, however, show that the all-pole model yields a sufficient filter to synthesize the PA of [s] sounds.

The samples which do not have the pole pair around 2kHz have been selected as the normal articulation of [s] sounds. The [ʃ], was never selected as an answer in the Exp. 1. We conclude that fricatives having a peak in the range of 1400 – 2400 Hz tend to be identified as the PA of [s]. The first order moment of the graph in Fig. 5 is approximately 1800 Hz, and at that frequency the graph has the highest score. This is comparable to the result of LP analysis in Section 3.

Because results may contain a bias due to the limited number of speech therapists used in Exp. 1 and Exp. 2, we will apply this method to an extended study which will use more speech therapists. Since the bandwidth of the pole pair around 2 kHz was fixed at 100 Hz in this study, we will also study the effects of variable bandwidth in the future.

Acknowledgements

We would like to acknowledge gratefully the assistance of the speech therapists who participated in the perceptual experiment in this study.

REFERENCES

[1] M. E. Morlay: *Cleft palate and speech*, ed. 7th, E. & S. Livingston, Edinburgh and London, pp. 173–210, 1970.

[2] K. Okazaki, M. Kato and T. Onizuka: "Palate morphology in children with cleft palate with palatalized articulation," *Ann. Plast. Surg.*, Vol. 26, pp. 156–162, 1991.

[3] M. Abe: "Nasopharyngeal articulation: A particular type of articulation error," *The Japan Journal of Logopedics and Phoniatrics*, Vol. 28, No. 3, pp. 239–256, July 1987.

[4] M. Kato: "Lateral articulation: Observation of articulatory movement using electro tooth-palatography," *The Japan Journal of Logopedics and Phoniatrics*, Vol. 32, No. 1, pp. 18–22, January 1991.

[5] K. Okazaki, T. Onizuka, M. Abe, et al.: "Palatalized Articulation as a Type of Cleft Palate Speech: Observation by Dynamic Palatograph and Cineradiograph," *The Japan Journal of Logopedics and Phoniatrics*, Vol. 21, No. 2, pp. 109–120, April 1980.

[6] D. R. Dickson: "An acoustic study of nasality," *J. Speech and Hearing Research*, Vol. 5, pp. 103–111, 1962.

[7] M. F. Schwartz: "Acoustic measures of nasalization and nazality," *Communicative Disorders related Cleft Lip and Palate* edited by K. R. Bzoch, pp. 194–200, Little Brown and Company, Boston, 1979.

[8] K. Okazaki and M. Kato: "Acoustic Analysis of Palatalized Articulation of [s] Sounds," *The Japan Journal of Logopedics and Phoniatrics*, Vol. 29, No. 3, pp. 225–231, July 1988.

[9] M. Wakumoto: "Quantitative Technique for Evaluation of Japanese Misarticulations — Acoustical Characteristics of Japanese Palatalized Articulation —," *Jpn. J. Cleft Palate Assoc.*, Vol. 14, pp. 21–43, 1989.

[10] T. Fukuda: "Acoustic Characteristics of Palatalized Articulation in Cleft Palate Children and Changes following Speech Therapy," *J. Osaka Univ. Dent. Soc.*, Vol. 38, pp. 8–35, 1993.

[11] Y. Ozawa and K. Okazaki: "Articulatory Tongue Movements and Acoustic Characteristics of Palatalized Articulation in Cleft Palate Patients," *The Japan Journal of Logopedics and Phoniatrics*, Vol. 35, pp. 322–330, 1994.

[12] Flanagan, J. L.: *Speech Analysis; Synthesis and Perception*, Springer-Verlag, Berlin, Germany, 1972.