Ann. Bull. RILP, No. 11, 85-91 (1977)

ON THE INFLUENCE OF CONTEXT UPON PERCEPTION OF VOICELESS FRICATIVE CONSONANTS*

Osamu Kunisaki** and Hiroya Fujisaki

Introduction

It is widely recognized that realization of a phoneme is influenced by its neighbouring phonemes in the process of speech production, so that its acoustic characteristics are generally dependent on the context. In order for the speech communication to be reliable, however, the context dependency must be corrected in the process of speech perception. In other words, speech perception must also be context-dependent to cope with the context-dependency of speech production. Examples of context dependency have already been reported often in vowel perception, [1] but less often in consonant perception.[2] In fact, few studies have ever been published on quantitative analysis of the context dependency in perception of fricative consonants.

The study to be reported here is a part of our efforts to gain quantitative knowledge concerning such context-dependencies both in production and in perception of speech, which is necessary both for automatic speech recognition and for speech synthesis.

Spectral Model of Voiceless Fricatives

Before describing our perceptual study, we will briefly summarize our analytical study of voiceless fricative consonants of Japanese. [3]

The excitation source of the voiceless dental fricative /s/ and the palatal fricative /š/ can be considered as the random pressure fluctuation caused by turbulence of the airflow at the outlet of the vocal tract constriction, formed by the tongue and the upper gum. Although the actual configuration of the vocal tract is fairly complex, the general characteristics of sounds produced by such a configuration can be understood on the basis of a simplified model, obtained by regarding the entire vocal tract as consisting of three parts, that is, the constriction, the front cavity and the back cavity as shown by the equivalent circuit of Fig. 1. Each of these three parts is approximated by an acoustic tube of uniform cross-sectional area.

In this configuration, the transfer admittance relating the turbulent noise pressure source to the volume current at the lips is found to possess a pole at the origin, and then one zero and two poles in the ascending order of frequency, if the frequency range is limited up to 5 kHz. On the other hand, frequency spectrum of the turbulent noise source is assumed to be fairly flat over the frequency range. The radiation transfer impedance can be approximately represented by an emphasis characteristics of 6 dB/oct.

---

* Paper presented at the IX International Congress on Acoustics, Madrid, July, 1977.

** Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo, Japan.

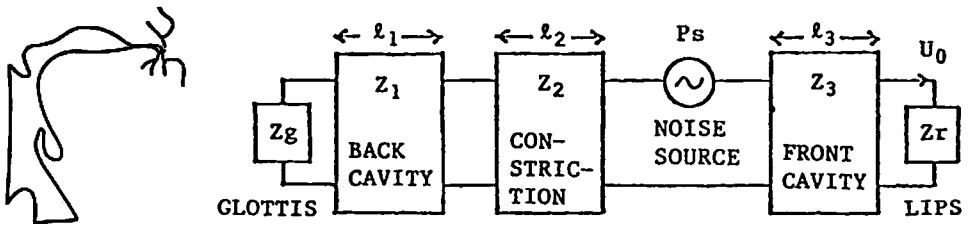

Fig. 1. Midsagittal section illustrating an articulatory configuration appropriate to the fricative /s/ (left) and equivalent circuit for the production mechanism of voiceless fricative consonants (right).

These considerations lead to the model for the actual frequency spectra of the two voiceless fricative consonants as shown in Equation 1.

$$P(s) = K\left[\frac{(s-s_{z1})(s-s_{z1}^*)}{s_{z1}s_{z1}^*}\right]\left[\frac{1}{s}\prod_{j=1}^{2}\frac{s_{pj}s_{pj}^*}{(s-s_{pj})(s-s_{pj}^*)}\right]s^{1+\alpha} \qquad (1)$$

$$s = 2\pi jf,\ s_{z1} = 2\pi(jF_{z1} - B_{z1}),\ s_{pj} = 2\pi(jF_{pj} - B_{pj})$$

In this equation, the first term K represents noise source intensity, the second and third terms respectively represent contributions of zero and poles within the frequency range of interest, and the last term represents the combined effect of radiation characteristics, higher poles and zeros, and deviation of the noise source spectrum from exactly flat characteristics. Figure 2 shows the approximate shape of the frequency spectrum given by such a model.

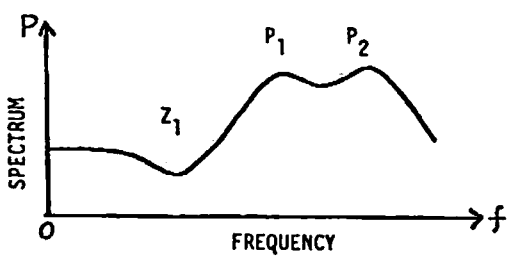

Fig. 2. Frequency characteristics of the spectral model.

The validity of the proposed model was tested by its ability to approximate the actual frequency spectra. The speech materials used for this purpose were 60 words of all possible combinations of the five Japanese vowels and the two voiceless fricative consonants, and were uttered by a male speaker.

The first step of the analysis was derivation of frequency spectra of consonantal segments of speech, extracted by a 50 msec hanning window. The frequency range from 0.3 to 5.0 kHz was then divided into 24 frequency

bands according to the Mel scale and the averaged level within each band was calculated to represent the smoothed spectral envelope. The second step of the analysis was determination of pole and zero frequencies as well as the overall spectral envelope of the model for each of these measured spectra by the method of Analysis-by-Synthesis, namely by finding their values that yield the best approximation to the measured spectral envelope in terms of the least mean squared error criterion. Figure 3 shows examples of such approximations of spectra of the dental and the palatal consonants followed by the vowel /a/. The smooth curves here indicate best approximations based on the model.

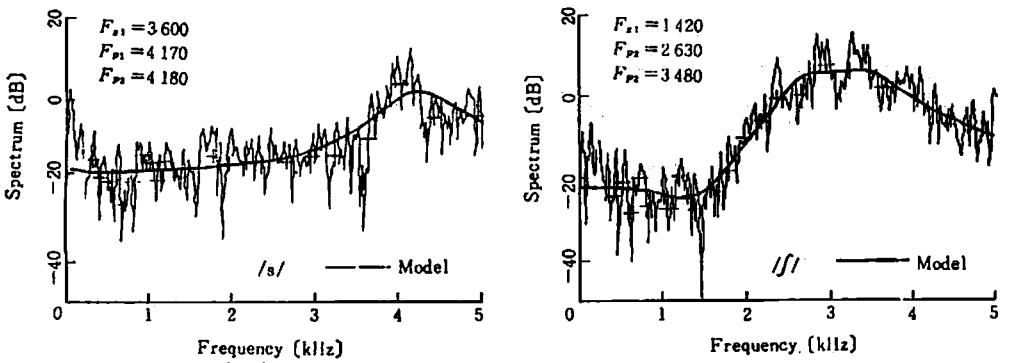

Fig. 3. Examples of parameter extraction of /s/ and /s̬/ by Analysis-by-Synthesis.

As expected, the extracted parameters of voiceless fricatives vary with the adjacent vowels, especially with the following vowels. Figure 4 shows the influence of the following vowels /a/ and /u/ on the two extracted parameters, that is, the frequency of the zero and that of the first pole. For example, the parameter values are smaller when the consonant is followed by /u/ than when it is followed by /a/. This fact in turn suggests that perception of these sounds may be influenced by context even if their physical parameters are fixed.

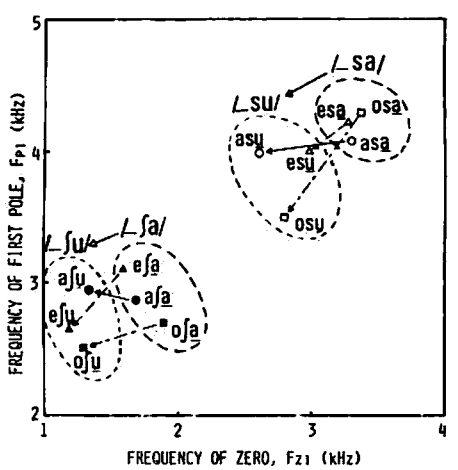

Fig. 4. Influence of following vowels /a/ and /u/ on extracted parameters of fricative consonants /s/ and /s̬/ in the word-medial position.

Experimental Techniques of Identification Tests

In order to confirm quantitatively these possibilities, the influence of context upon perception of voiceless fricatives was examined by identification tests of synthetic stimuli.[4] Synthetic vowel-consonant-vowel sounds were generated by computer simulation of a terminal-analog speech synthesizer, and were used in identification tests to determine the phoneme boundary between the dental consonant /s/ and the palatal consonant /$\check{s}$/.

Figure 5 shows the block diagram of the synthesizer consisting of two parts, that is the vowel part and the consonant part. The vowel synthesizer consists of a buzz source followed by five cascated poles and an emphasis characteristic which represents the effect of radiation. The consonant synthesizer consists of a random noise generator followed by one zero and two pole characteristics.

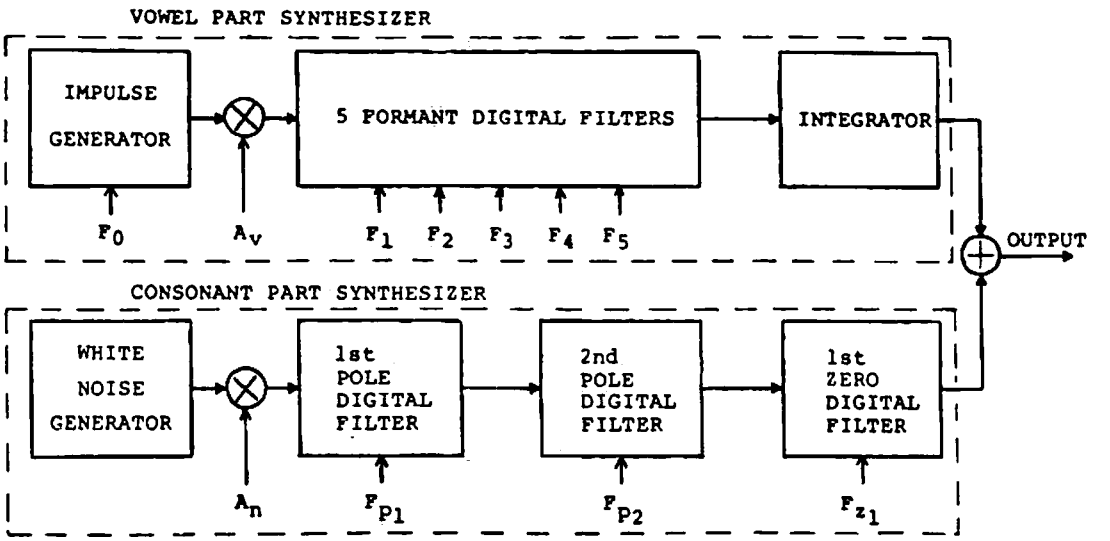

Fig. 5. Block diagram of synthesis of vowel-consonant-vowel stimuli.

Based on the analysis of measured spectra of voiceless fricatives in /sa/ and /$\check{s}$a/ syllables, parameters of the consonantal parts of synthetic stimuli were selected at points in the parameter space such that they represented the two typical sounds as well as intermediate sounds. As shown in Fig. 6, a set of ten points were selected over the entire range, represented by a straight line in the three-dimensional parameter space constructed by frequencies of the zero and the two poles.

Figure 7 shows the time chart of vowel formant frequency and source intensity control for the synthesis of vowel-consonant-vowel. The formant patterns used for the vowel part were obtained by simplifying those found in utterances of a male speaker, where the first and second formant frequencies were expressed by a set of straight lines, while higher formant frequencies were held constant as shown in Fig. 7. The fundamental frequency of the vowel was held constant at 150 Hz, and the intensity of the stationary part of the consonant relative to that of the stationary part of the vowel was fixed at -18 dB, based on preliminary analysis and listening tests.

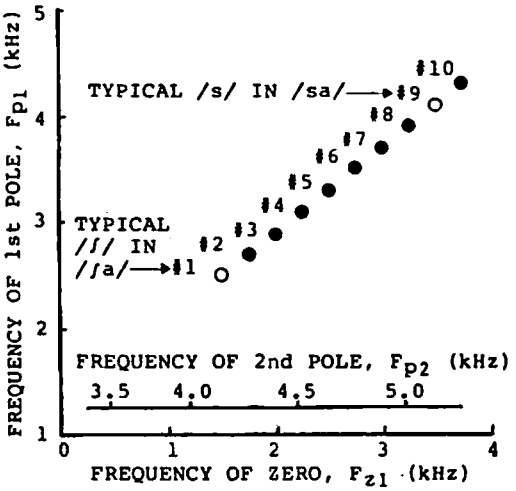

Fig. 6. Parameters of synthetic voiceless fricative consonants used as stimuli for identification tests.

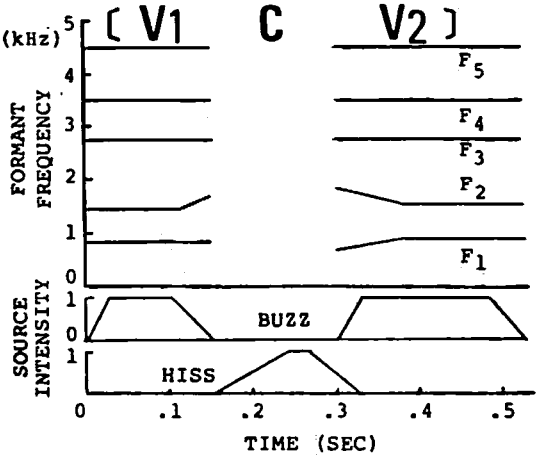

Fig. 7. Time chart of vowel formant frequency and source intensity control for synthesis of VCV stimuli.

Using these synthetic stimuli, two sets of identification tests were performed: The first set was designed to examine the effect of the initial vowel upon the perception of the intervocalic consonant in vowel-consonant-vowel type words, where the final vowel was always fixed at /a/. The second set of identification tests was designed to examine the effect of the final vowel on consonantal perception in the same type of words, but the initial vowel was fixed at /a/. In each of these identification tests, the ten different stimuli were arranged in random order and presented to subjects ten times at intervals of 4 sec. The subjects were three male adults with normal hearing.

Results of Identification Tests

Figure 8 shows examples of results of identification tests plotted on a normal paper. The abscissa indicates the stimulus number along the frequency of the spectral zero, the ordinate indicates the percentage of /s/-judgments. Empty circles and filled circles respectively correspond to tests with /u/ and /e/ as the final vowel. The straight lines indicate estimated normal distributions determined by the method of the least-mean squared error weighted by Müller-Urban coefficients. The phoneme boundary between the two fricatives can be defined as the point corresponding to 50 percent /s/-judgments.

Figure 9 shows the influence of the initial vowel on the phoneme boundary of intervocalic fricatives in VCV words. We can see that the intervocalic consonant is slightly more often identified as palatal when it follows back vowels /o/ or /a/ than when it follows non-back vowels /e/, /i/, or /u/, while the influence of the initial vowel on the phoneme boundary is rather small.

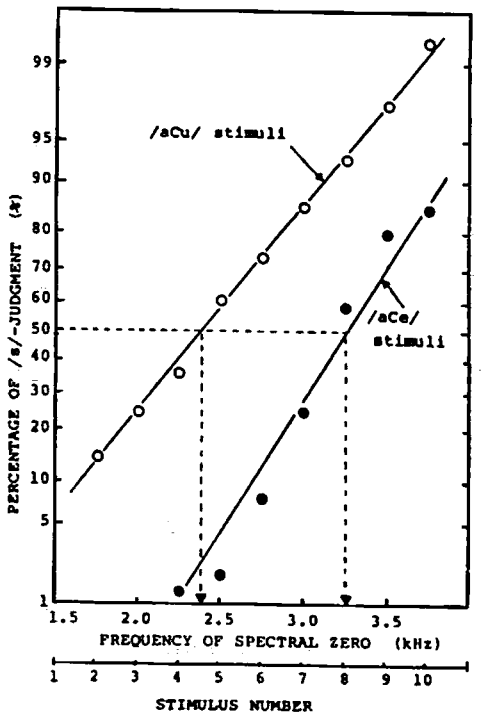

Fig. 8. Examples of results of identification test for voiceless fricative consonants in VCV stimuli /aCu/ and /aCe/, and their approximation in terms of normal distribution with Müller-Urban weighting.

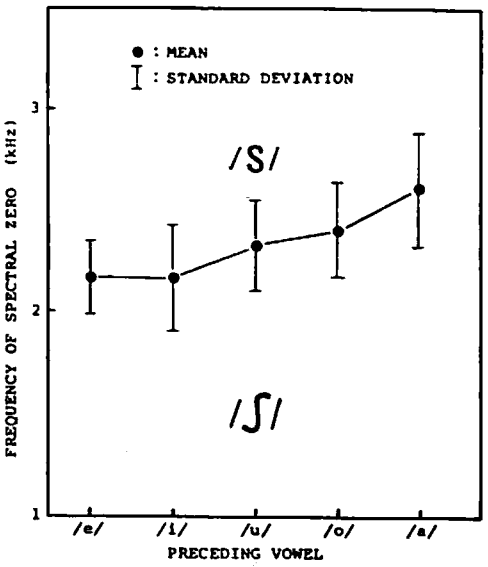

Fig. 9. Influence of the preceding vowel on phoneme boundary between /s/ and /ʃ/ in /VCa/ stimuli.

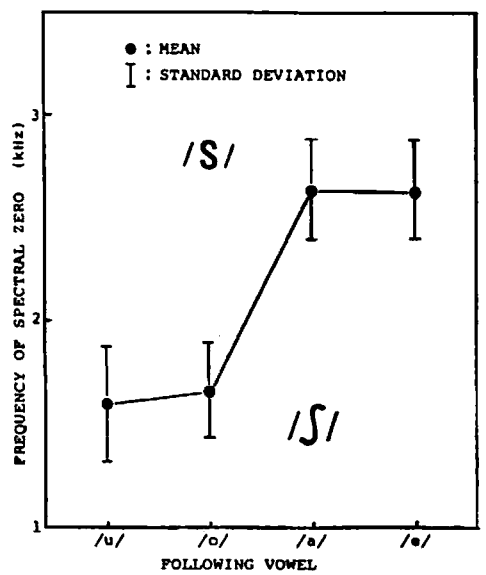

Figure 10 shows the influence of the final vowel on the phoneme boundary of intervocalic fricatives. Compared with the previous figure, the influence of the final vowel on consonantal identification is found to be fairly large. We can see that the intervocalic consonant is significantly more often identified as palatal when it is followed by /a/ or /e/ than when it is followed by /u/ or /o/.

Fig. 10. Influence of the following vowel on phoneme boundary between /s/ and /ʃ/ in /aCV/ stimuli.

These results suggest that the effects of lip articulation of the final vowel are reflected on the perception of the preceding consonant. That is to say, the lip rounding in /u/ and /o/ as final vowels tends to decrease the frequencies of spectral poles and zero of the preceding fricative consonants, and shifts the phoneme boundary in the same direction.

Concluding Remarks

The influence of vowel context upon neighbouring voiceless fricative consonants /s/ and /š/ has been examined both from the point of view of production and from that of perception. Analysis of these consonants in terms of a spectral model has revealed a predominant influence of the vowel that immediately follows the consonant. Identification tests using synthetic vowel-consonant-vowel stimuli have generally confirmed the results of acoustic analysis. In particular, these experiments have indicated that a large difference exists between influence of vowels /u/ and /o/ and that of vowels /a/ and /e/ upon the phoneme boundary between /s/ and /š/, suggesting that the influence of context in speech production is corrected in speech perception. These findings are of significance both for speech synthesis and for automatic speech recognition.

References

1. Strange, W., R. R. Verbrugge, D. P. Shankweiler and T. R. Edman, (1976), "Consonant Environment Specifies Vowel Identity," Haskins Laboratories Status Report on Speech Research, SR-45/46, 37-57.

2. Hasegawa, A. and R. G. Daniloff (1976), "Effects of Vowel Context upon Labelling the /s/-/š/ Continuum," 91st ASA Meeting, Paper M10.

3. Fujisaki, H. and O. Kunisaki (1976), "Analysis, Recognition and Perception of Voiceless Fricative Consonants in Japanese," Rec. 1976 IEEE ICASSP, 158-161. [Also in Ann. Bull. RILP, No. 10, 145-156.]

4. Kunisaki, O, T. Matsuo, and H. Fujisaki (1976), "Perceptual Study of Voiceless Fricative Consonants Using Synthetic Stimuli," Rec. Spring Meeting, Acoust. Soc. Japan, 327-328.