Ann. Bull. RILP, No. 9, 123-126 (1975)

ANALYSIS AND RECOGNITION OF VOICELESS FRICATIVE CONSONANTS IN JAPANESE

Hiroya Fujisaki and Osamu Kunisaki*

Accurate extraction of the acoustic features and their effective utilization are essential in the automatic recognition of speech. The present report describes an approach to the analysis and recognition of the Japanese voiceless fricative consonants /s/ and /ʃ/, on which little has been published as compared with vowels and voiced consonants. [1-3] Based on an equivalent circuit representation of the production mechanism of these consonants, a model is derived for their frequency spectra up to 5 kHz. The model is then utilized to extract the acoustic parameters from measured spectra of these consonants, and the extracted parameters are shown to be effective in their automatic recognition. Various simplifications of the model are also investigated.

A Model for Frequency Spectra of Voiceless Fricative Consonants

The excitation source of the voiceless fricative consonants /s/ and /ʃ/ is the random pressure fluctuation caused by turbulence of the airflow at the outlet of the constriction of the vocal tract. [4]

If we regard the entire vocal tract as consisting of three parts, viz., the constriction, the front cavity and the back cavity, and assume that each of them can be approximated by a lossless acoustic tube of uniform cross-sectional area, we obtain the equivalent circuit of Fig. 1. By assuming

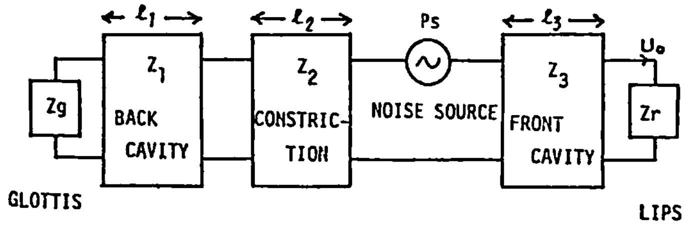

Fig. 1: Equivalent circuit for the production mechanism of voiceless fricative consonants.

further that the input impedance of the back cavity is negligible as compared to that of the constriction, and that the radiation impedance from the lips is also negligibly low, the transfer admittance T relating the turbulent noise pressure source and the volume current at the lips can be given by

$$T = \frac{-j\cdot\cos 2\pi f \ell_2/c}{Z_2 \sin 2\pi f \ell_2/c \cdot \cos 2\pi f \ell_3/c + Z_3 \sin 2\pi f \ell_3/c \cdot \cos 2\pi f \ell_2/c} \quad (1)$$

* Dept. of Electrical Engineering, Faculty of Engineering, Univ. of Tokyo.

where ℓi and Zi respectively denote the length and the characteristic impedance of the i-th section of the vocal tract, and the suffixes 2 and 3 respectively refer to the constriction and the front cavity. The transfer admittance is found to possess a pole at the origin, and then one zero and two poles in the ascending order of the frequency, if the ratio $\ell 2/\ell 3$ lies between 1 and 3, which is the case for both /s/ and /ʃ/. These considerations, based on the simplified equivalent circuit of Fig. 1, lead to the following model for the actual frequency spectra of /s/ and /ʃ/ :

$$P(S) = K \frac{(S-S_{z1})(S-S_{z1}^{*})}{S_{z1}S_{z1}^{*}} \cdot \frac{1}{S} \prod_{i=1}^{2} \frac{S_{pi}S_{pi}^{*}}{(S-S_{pi})(S-S_{pi}^{*})} \cdot S^{1+\alpha} \quad (2)$$

$$S=2\pi jf \ , \ S_{z1}=2\pi(jF_{z1}-B_{z1}) \ , \ S_{pi}=2\pi(jF_{pi}-B_{pi})$$

where the last term in Eq. (2) represents the combined effect of the radiation characteristics, higher poles and zeros, and the deviation of the source spectrum from exactly flat characteristics. [5]

Analysis of Voiceless Fricative Consonants [6]

The validity of the proposed model can be tested by its ability to approximate the actual frequency spectra of voiceless fricative consonants. The speech materials used for this purpose were 60 words of CV and VCV type consisting of all possible combinations of the five Japanese vowels and the two consonants /s/ and /ʃ/, and were uttered by a male speaker. These materials were sampled at 20 kHz with an accuracy of 11 bits/sample, and the logarithmic frequency spectrum was calculated for a 50 msec segment taken out of each of the consonants.

The frequency range from 0.3 to 5.0 kHz was then divided into 24 frequency bands according to Mel scale and the averaged level within each band was calculated to represent the smoothed spectral envelope. For each of these measured spectra, the best approximation in terms of the least mean squared error criterion was obtained by finding the frequencies of poles and zero as well as the overall spectral slope of the model, while the Q's of poles and zero were held constant at their respective values found in preliminary analysis.

Examples of such an approximation are shown in Fig. 2 for /s/ and /ʃ/, together with parameter values thereby extracted. Results of similar approximations by other models with reduced numbers of poles and zeros indicate that models incorporating a zero are superior to those without zero in the approximation of /ʃ/-spectra, while the differences are less for /s/-spectra.

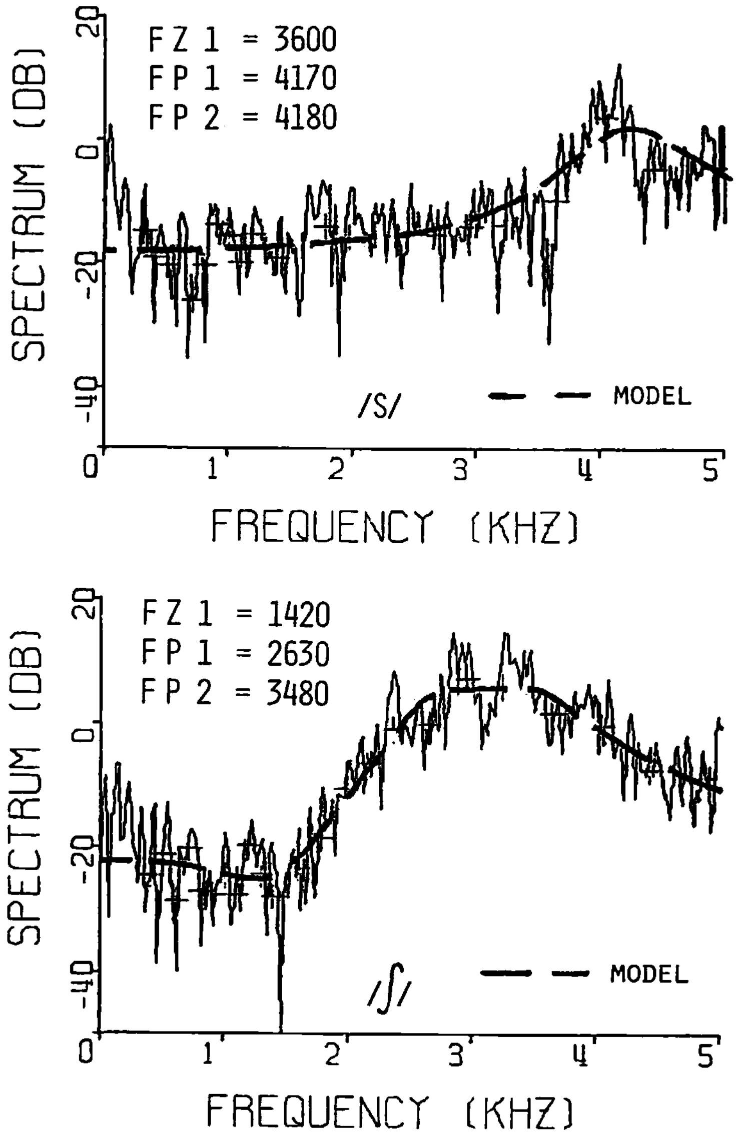

Fig. 2: Examples of parameter extraction of /s/ and /ʃ/ by Analysis-by-Synthesis.

Recognition of /s/ and /ʃ/ in the Parameter Space

The close agreement between the model and the measured spectra of /s/ and /ʃ/ indicates the model's capability of expressing the acoustic characteristics of these sounds, and suggests the use of extracted parameters for their automatic recognition. In order to evaluate the model from this point of view, the optimum linear discriminant function was determined in the three-dimensional feature space (Fz_1, Fp_1, Fp_2) to minimize the probability of error in the classification of /s/- and /ʃ/-samples, and was used to recognize each sample as either /s/ or /ʃ/. For the 60 samples, complete separation of /s/ and /ʃ/ was possible by this method even without Fp_2. Models with reduced numbers of poles and zeros were also evaluated in the same way, and the results are summarized in Table 1.

These results indicate the importance of the zero frequency (Fz_1) for the correct recognition of fricative consonants, and validate the use of the proposed spectral model for its extraction.

Models	No. of zeros	1	1	0	0
	No. of poles	2	1	2	1
Recognition rate (%)		100	100	97	92

Table 1: Comparison of spectral models for voiceless fricative consonants in the automatic recognition of /s/ and /ʃ/.

References

1. Hughes, G. W. and M. Halle (1956) "Spectral Properties of Fricative Consonants," J. Acoust. Soc. Am., 28, 303-308.
2. Heinz, J. M. and K. N. Stevens (1961) "On the Properties of Voiceless Fricative Consonants,": J. Acoust. Soc. Am., 33, 589-596.
3. Heinz, J. M., (1961) "Analysis of Fricative Consonants," MIT R. L. E. QPR No. 60, 181-184.
4. Heinz, J. M., (1958) "Model Studies of the Production of Fricative Consonants," MIT R. L. E. QPR No. 50, 146.
5. Stevens, K. N. (1971) "Airflow and Turbulence Noise for Fricative Consonants: Static Considerations," J. Acoust. Soc. Am., 50, 1180-1192.
6. Fujisaki, H. and O. Kunisaki (1975) "Parameter Extraction for Automatic Recognition of Voiceless Fricative Consonants in Japanese," Nat. Conv. Rec., Inst. Elect. Comm. Eng. Japan., 1240.