ANN.Bull.RILP(1990)No.24, 89-95

PERCEPTUAL FREQUENCY NORMALIZATION OF FREQUENCY-COMPRESSED OR EXPANDED VOICELESS STOPS

Sotaro Sekimoto

1. Introduction

The efficiency of hearing aids with frequency-lowering has been studied[1]. The frequency axis of speech was compressed downward by a PARCOR speech analysis-synthesis method. It was observed in a hearing test for this speech that considerable improvement in articulation scores was obtained for vowels, whereas little improvement was seen for consonants, especially for voiceless consonants. For vowels, it was found in a subsequent study that speech in which the frequency axis was compressed or expanded was identified as unaltered over a wide range of frequency expansion or compression ratio. It was thus concluded that a similar perceptual normalization to that observed for ordinary male, female and children's voices plays an important role in identifying frequency-compressed speech, and that this produced the improvement in the articulation scores in the hearing test described above. For consonants, on the other hand, although it is supposed that perceptual normalization again plays a role in the perception of frequency-compressed consonants, few studies have been made on the normalization of consonantal speech[2].

In our previous report[3], it was reported that the expansion or compression of the frequency axis was not normalized perceptually in word-initial voiceless fricative consonants. In the present paper, the results of perceptual experiments are described to clarify the characteristics of perceptual normalization for the expansion or compression of a frequency axis in word-initial voiceless stop consonants.

2. Method

One result of our previous experiment[3] was that the identification of word-initial voiceless fricative consonants, whose axes were compressed or expanded, was based on the absolute pole frequency of the prevocalic fricative noise. The reason for this was assumed to be that the fricative noise used was composed of only a single component, so that a relative comparison between two or more components could not be done within the noise itself. In this experiment, therefore, two types of noise with different noise component structures were used. From a preliminary experiment, it was determined that the above two kinds of noise, that is, single-pole noise and multiple-pole noise, could be adopted as the prevocalic noise for a word-initial synthetic voiceless stop consonant. A block diagram of the synthesizer is shown in Fig. 1. The single-pole noise was synthesized using the lower branch of the diagram with attenuator "Af" switched to "On". The multiple-pole noise was synthesized using the upper branch where the attenuator "Av" was "Off" and "Aa" was "On". The former was

the condition in which absolute identification was expected, and the latter was that where relative identification was expected.

The onset frequency of the resonant circuit, F_p for single-pole noise and F_2 for multiple-pole noise, was systematically varied and subjected to a hearing test to determine the perceptual phoneme boundary between /t/ and /k/. The phoneme boundaries were compared for various ratios of frequency-compression or expansion. The vowel was /a/.

Stimuli

Stimuli were synthesized with a software terminal-analog speech synthesizer. Rosenberg's C-waveform was used as a voice source.[4] The time patterns of the resonant frequencies of the noise poles are shown in Fig. 2. The initial 10msec of the noise period was kept stable to simulate a noise burst, and the following 55msec was a transition period simulating aspiration. The transition pattern was approximated by a step-response of a 1st-order linear system. The frequency (f_n) at t was defined as

$$f_n(t) = F_{ne} - F_{ns} \exp(-t/T_c)$$

where, F_{ne}, F_{ns}, T_c were the target frequency of the following vowel, the extent of the frequency transition and the time-constant of the transition, respectively. The time-constant was always 10msec. The value of the second formant frequency of the vocalic portion was used as a target. The values of the formant frequencies of the vocalic portion F_1, F_2, F_3, F_4, and F_5 were held constant at 800Hz, 1200Hz, 2400Hz, 3500Hz, and 4500Hz, respectively. The fundamental frequency was changed from 114Hz to 80Hz in the initial 300 msec period, then held constant until 400msec. The same fundamental frequency pattern was used for each frequency-compression or expansion ratio.

The compression or expansion of the frequency axis was accomplished by lowering or raising the sampling frequency of the filters of the synthesizer. The sampling frequency for the uncompressed condition was 20 kHz. The speech stimuli were produced on a software synthesizer and D/A-converted at 12-bit precision. The signal was then low-pass filtered with a cutoff of -135 dB/oct and recorded on a DAT (Digital Audio Tape). The cutoff frequency of the low-pass filter was dynamically changed in proportion to the sampling frequency by a factor of 0.45.

Experimental parameters

(1) The onset resonant frequencies of the noise pole (F_p for the single-pole noise and F_2 for the multiple-pole noise) : 1100Hz, 1300Hz, 1500Hz, 1700Hz, 1900Hz, 2100Hz, 2300Hz.
(2) The frequency-compression or expansion ratios, which were defined as percent ratios of the frequency-compression or expansion against the uncompressed condition (100%) : 60% and 80% (compressed), 100% (uncompressed), 120% and 140% (expanded).

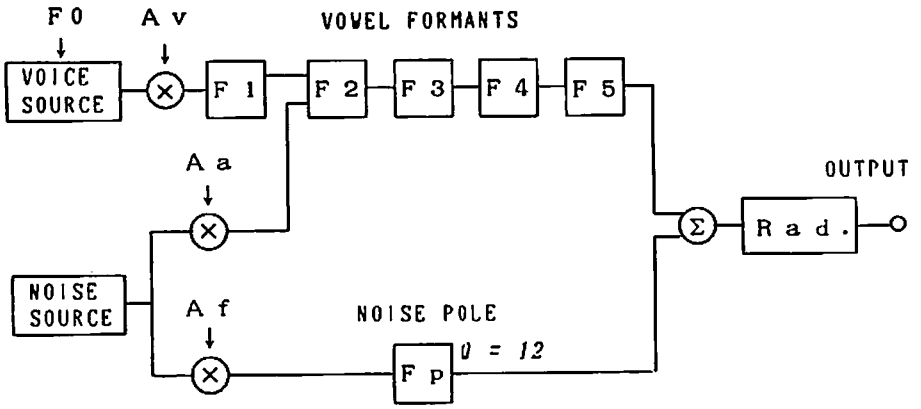

Fig. 1. A block diagram of the software speech synthesizer.

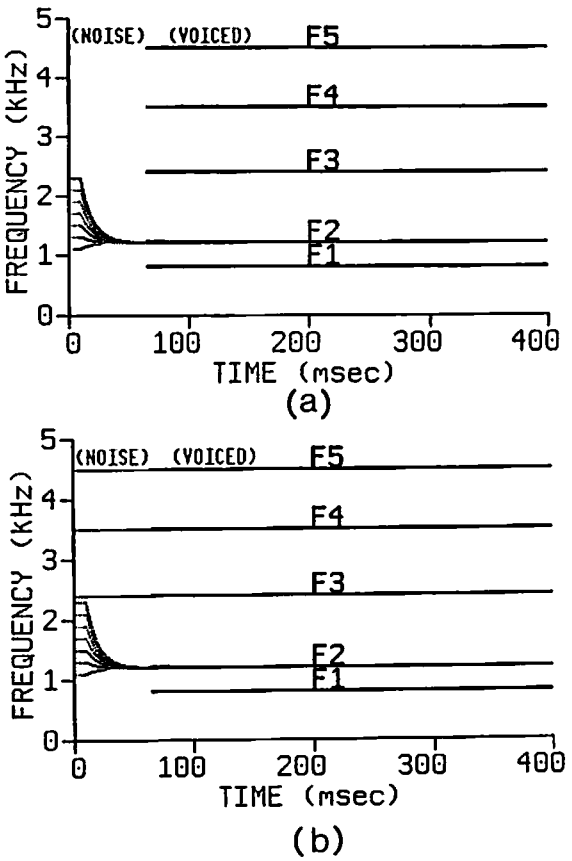

Fig. 2. The time patterns of the control parameters for synthesizing /ta/-/ka/ for the single-pole noise (Fig. a) and the multiple-pole noise (Fig. b).

Procedure

The stimuli were presented through binaural headphones (STAX SR-Lambda Signature) in a soundproof room at a comfortable presentation level (about 75 dB SPL). The hearing test was carried out with the constant method. The speech samples were presented in a random order. Subjects were requested to identify the synthetic stimuli as one of the Japanese voiceless stop consonants or voiceless fricative consonants. Three adult subjects participated.

3. Results and discussion

Results are shown in Fig. 3. The boundary pole frequencies, where the responses of /t/ and /k/ cross in the figure, shifted depending on the frequency-compression or expansion for both noise conditions, the single-pole noise and the multiple-pole noise. The inclination of the shift against the frequency-compression or expansion ratio was steeper for the multiple-pole noise than for the single-pole noise. However, the difference in the inclination seems to be insignificant with regard to the normalization. Rather, it reflects the role of the higher noise formants in the multiple-pole noise condition, which just act as an "weight" on the frequency component constituting a gross spectral shape which indicates a cue for identifying voiceless stop consonants.

From this experiment, it is clear that the noise structure of the prevocalic voiceless part does not contribute to the perceptual normalization for the compression or expansion of the frequency axis. Then, what is the cue of the perceptual frequency normalization? One candidate is the formant transition in the vocalic portion, but no transition was present in the stimuli used in this experiment. LaRiviere, et. al. reported for ordinary speech that the vocalic transition is neither a sufficient nor necessary cue for the recognition of initial voiceless stop consonants in prevocalic environments[5]. Another candidate is the transition in the prevocalic noise portion. In order to assess the contribution of the transition in the prevocalic noise portion, an additional experiment was carried out. Stimuli were synthesized so that the transition period was replaced by a silent gap, and were subjected to the listening test. The results are shown in Fig. 4. The boundary pole frequencies between /t/ and /k/ shift similarly with the variation in the frequency-compression or expansion ratios as in Fig. 3. It is apparent that the transition in the prevocalic portion does not contribute to the perceptual normalization of the frequency axis.

These results suggest that the cue for the perceptual normalization of the frequency axis in voiceless stop consonants can only be in some relationship between the noise burst 10msec in length and the stationary vocalic portion. Further study is clearly needed.

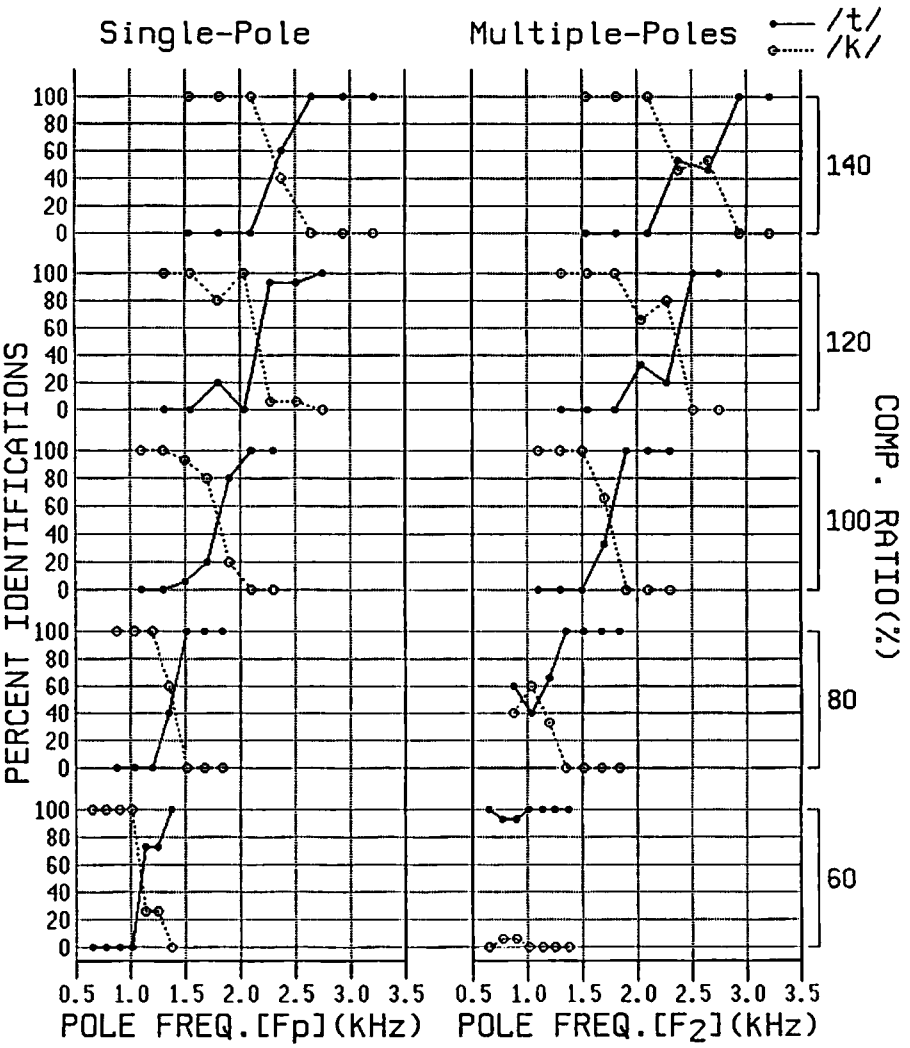

Fig. 3. The identification rates for /t/ and /k/ where the prevocalic noise is approximated by the single-pole noise (Left) and the multiple-pole noise (Right). The abscissa shows the absolute noise pole frequency after frequency-compression or expansion. The ordinate shows the identification rates for the various frequency-compression or expansion ratios. The identification rates for /t/ and /k/ are shown by solid lines and broken lines, respectively.

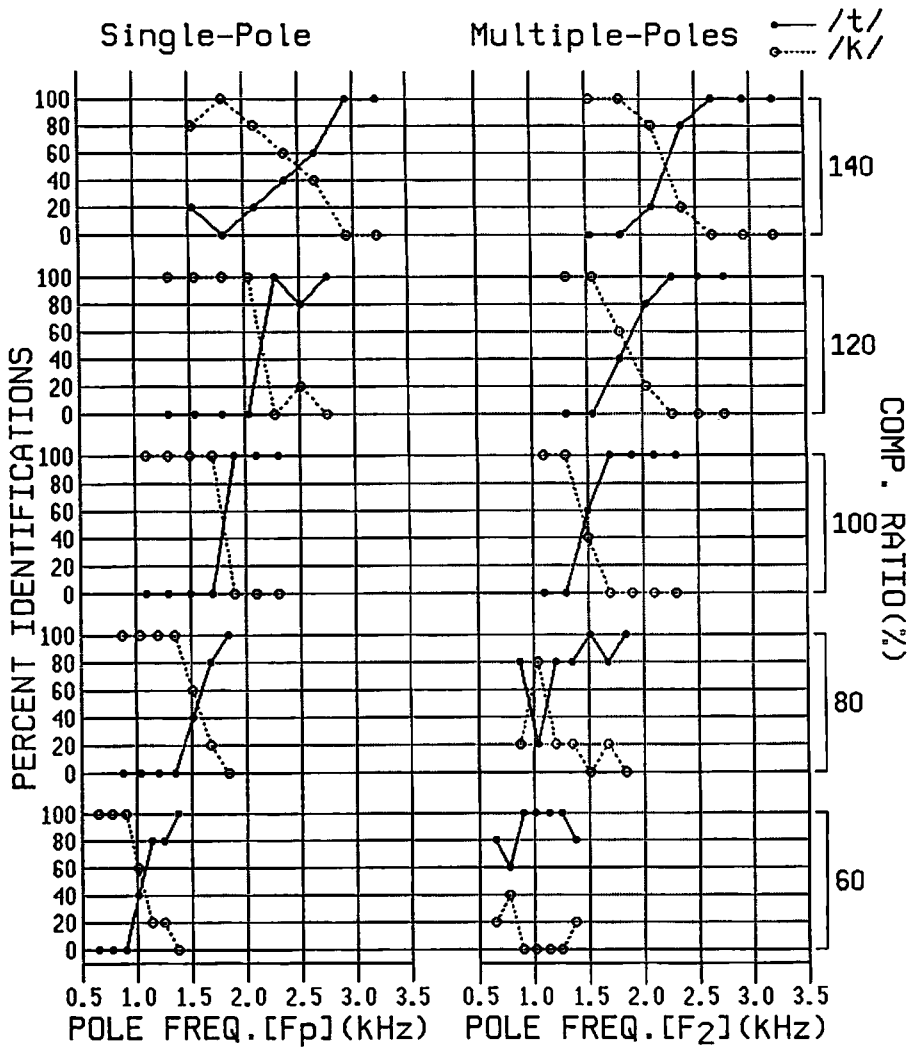

Fig. 4. The identification rates for /t/ and /k/ when the transition in the prevocalic portion was replaced by a silent gap.

4. Conclusion

The phonetic category boundary between /t/ and /k/ on the continuum of the noise-pole frequency shifted in relation to frequency-compression and expansion independently of the number of prevocalic noise poles. This result suggests that normalization for the compression or expansion of the frequency axis occurs in the identification of voiceless stop consonants. An effect of the transition in the prevocalic noise portion on the normalization was not observed.

References

1) Sekimoto, S., S. Kiritani, and S. Saito (1980); Intelligibility of frequency-compressed speech in low-pass filtered condition, Ann. Bull. RILP, 14, 181-193.
2) Sekimoto, S. (1982); Perceptual normalization of frequency scale, Ann. Bull. RILP, 16, 95-101.
3) Sekimoto, S. (1989); Normalization of frequency-compressed voiceless fricatives, Ann. Bull. RILP, 23, 39-49.
4) Rosenberg, A. E. (1971); Effect of glottal pulse shape on the quality of natural vowels, JASA, 49, 2(Pt. 2), 583-590.
5) LaRiviere, C., H. Winitz, and E. Herriman (1975); Vocalic transitions in the perception of voiceless initial stops, J. Acoust. Soc. Am. 57, 470-475.