Ann.Bull.RILP(1985)No.19, 75-83

PHONEME BOUNDARY SHIFT IN THE IDENTIFICATION OF FREQUENCY EXPANDED VOWELS

Sotaro Sekimoto

1. Introduction

In a previous study by the present author, it was observed that frequency expanded vowels, whose frequency axes were linearly compressed or expanded, were correctly identified as their originals within a range of frequency compression or expansion from 0.5 to 2.0 (Sekimoto, 1982). It is believed that some perceptual mechanism functions to normalize the variation of the frequency scale in perceiving such speech in which the frequency axis is compressed or expanded. In order to elucidate the perceptual mechanism for this normalization, perceptual experimental studies have been made (Fujisaki and Kawashima, 1970; Sekimoto, 1982; Sekimoto, 1983).

In the present study, the effects of the frequency expansion of the formant envelope, as well as the fundamental frequency, on the perceptual phoneme boundary in the F_1-F_2 plane in identifying frequency-expanded synthetic vowels were investigated.

2. Method

Experimental conditions

The identification test was made on frequency expanded synthetic vowels in which the first and the second formant frequencies (abbreviated as F_1 and F_2, respectively) were systematically varied. Five steady-state and isolated Japanese vowels uttered by an adult male speaker were used as speech materials. The conditions of the frequency expansion ratio and the fundamental frequency ratio are shown in Table 1. There, the frequency expansion ratio and the fundamental frequency ratio are defined as ratios of the frequency range of the frequency expanded speech and the fundamental frequency to those of the original vowels, respectively. The condition of the frequency expansion ratio was combined with one of the fundamental frequency ratios in the same file. The condition denoted as N indicates that of noise excitation.

Table 1 Conditions of the frequency expansion ratio and the fundamental frequency ratio

FREQUENCY EXPANSION RATIO	1.0	1.4	2.0
FUNDAMENTAL FREQUENCY RATIO	1.0, N	1.0, 1.2, 1.4, N	1.0, 1.4, 2.0, N

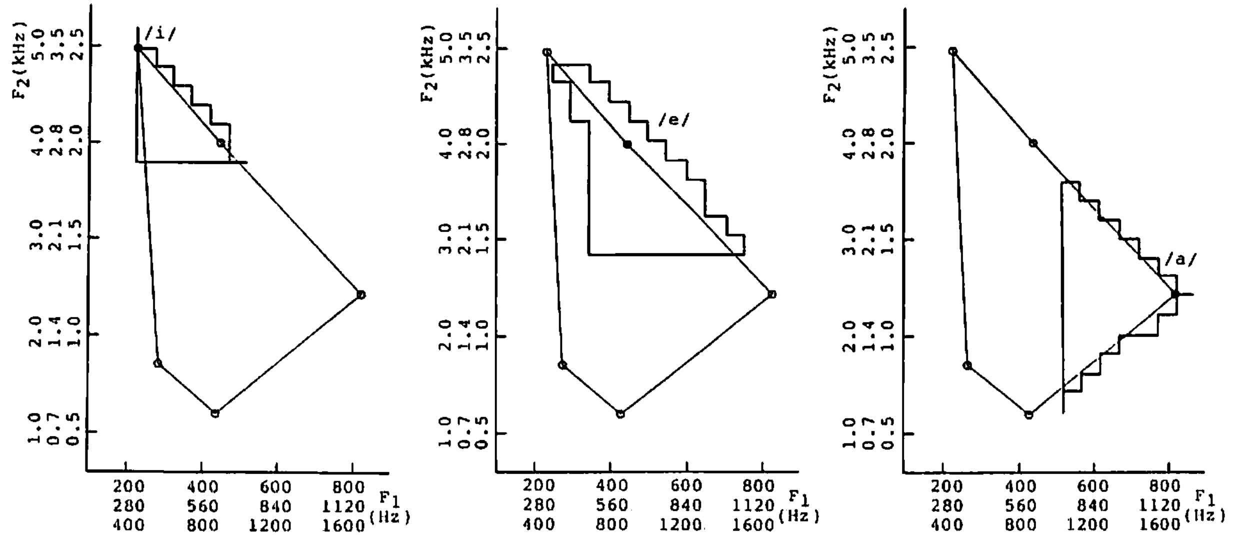
The first and the second formant frequencies were varied around the typical values for each vowel within the range on the F_1-F_2 plane shown in Fig. 1, in 50 Hz and 100 Hz steps, respectively, if the frequency axis was not expanded. These values were multiplied by a factor of the value same as the frequency expansion ratio when the frequency axis was expanded.

Procedure

The method of adjustment was adopted in the experiment. The experiment was made independently for the various combinations of the conditions shown in Table 1 and for the various vowels. The experiments were interactively made with the on-line computer. The variations in F_1 and F_2 were pointed with a stylus pen on a digitizer tablet connected to the computer. The deviation in F_1 and F_2 from the typical values of the first and the second formant frequencies of a vowel was entered as the displacement in the X- and Y- directions, respectively, from the center of the digitizer tablet. When the new value was entered, the frequency expanded speech with new F_1 and F_2 was immediately synthesized in real-time and presented to the subject. The stimuli were identified as one of the five Japanese vowels under forced choice by the subject. The response was entered in the computer with a stylus pen by pointing to the vowel menu at the top of the tablet. The stimuli were presented to both ears through headphones at a level of 75 dB SPL. An adult male subject participated.

Stimuli

The calculation process for the frequency expansion is shown in Fig. 2. The original speech signal was sampled and digitized at 10 kHz with a 12-bit accuracy. The digitized signal was subjected to PARCOR analysis to extract the LPC coefficients (a_0-a_{12}) and the fundamental frequency, through Hamming window 30 msec in length. The analysis was repeated at 5 msec intervals. Formant frequencies (F_1-F_6) and bandwidths (B_1-B_6) were calculated by solving the polynomial of the LPC coefficients. Extracted parameters so far: the fundamental frequency, the amplitude, the formant frequency and the formant bandwidths were stored on a disk. When the new values of F_1 and F_2 were entered, the formant frequencies were modified and used to compute new PARCOR coefficients by defactorization and a step-down recursion process. The new PARCOR coefficients and the other parameters on the disk were sent to the synthesizer to generate the frequency expanded speech. The frequency expansion was achieved by changing the sampling frequency at the synthesizer output. The speech processing was carried out on a VAX-11/780 computer with a floating point processor FPS-100. The calculation was made mainly on the FPS-100. The synthesizer program developed by the author ran 10 times faster than the real-time.



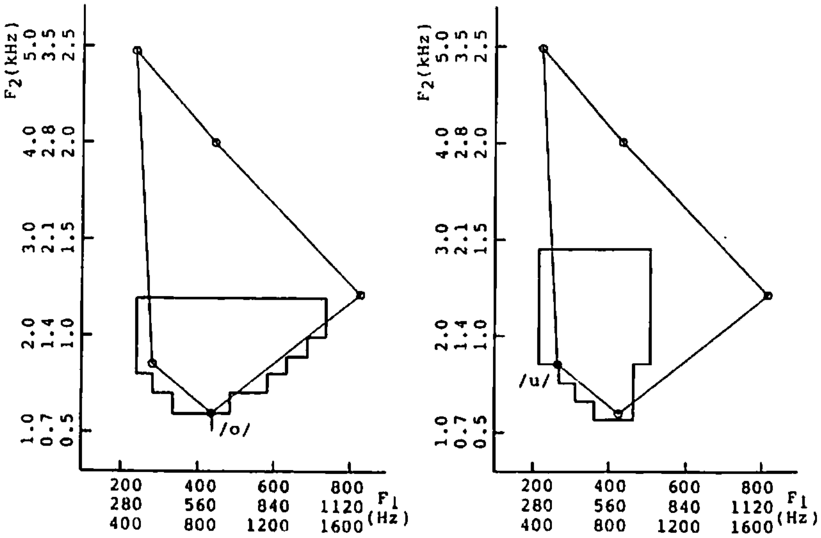


Fig. 1 The maximum ranges of the first and the second formant frequency variations. The ranges were expanded in proportion to the frequency expansion ratios when the frequency axis was expanded.

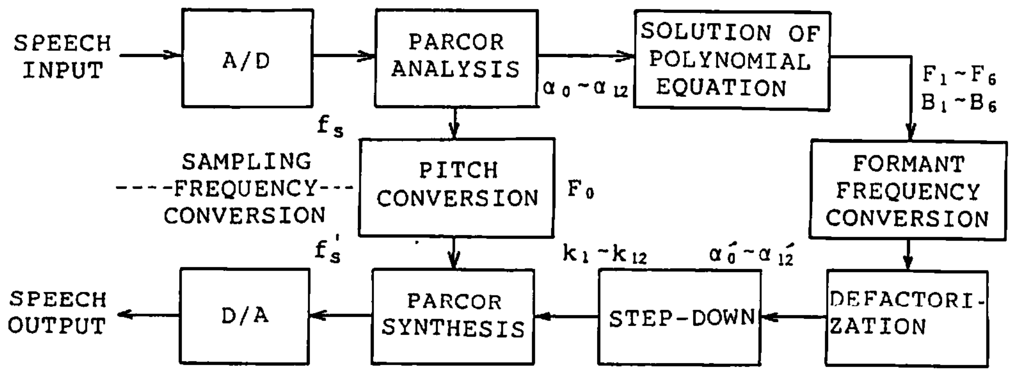


Fig. 2 The calculation process for the speech processing.

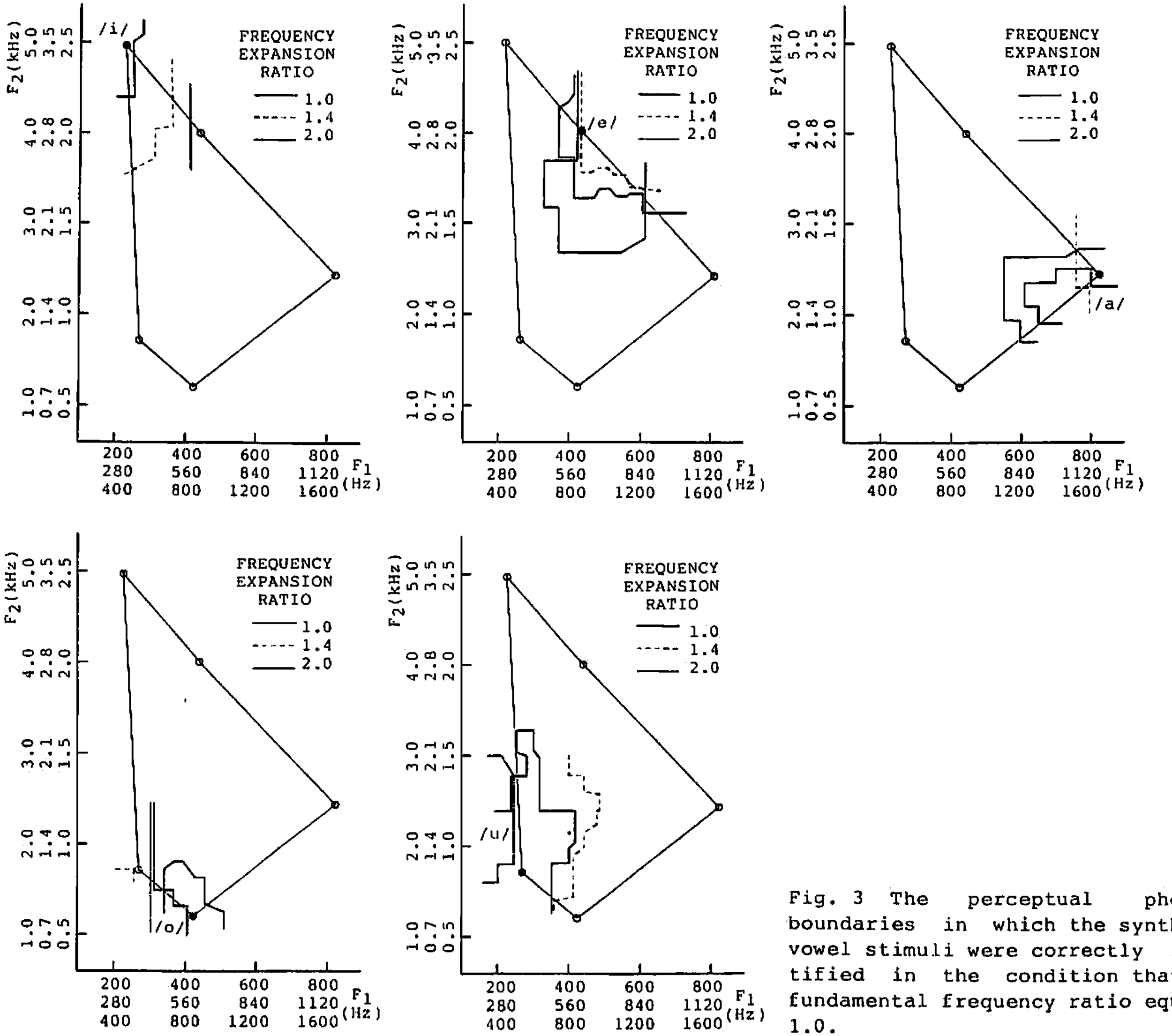


Fig. 3 The perceptual phoneme boundaries in which the synthetic vowel stimuli were correctly identified in the condition that the fundamental frequency ratio equaled 1.0.

3. Results and remarks

The perceptual phoneme boundary, or the ranges of F_1 and F_2 where the vowel was correctly identified, for different frequency expansion ratios are shown in Figs. 3, 4, and 5. Since the ranges of F_1 and F_2 were expanded in proportion to the frequency expansion ratios, the results for the different frequency expansion ratios are superimposed on the same figure so that the center frequencies of a vowel for different frequency expansion ratios are mapped onto the same point with different scales for F_1 and F_2.

The results when the fundamental frequency ratio equaled 1.0 are shown in Fig. 3. It was observed that the perceptual phoneme boundary shifted toward the outside of the center frequencies of each vowel as the frequency expansion ratio became greater for every vowels except /e/. The shift was remarkable for /a/ and /u/ when the frequency expansion ratio was 2.0, and for /o/ when the frequency expansion ratio was 1.4 and 2.0, where the original center frequencies were not contained within the phoneme boundary. As for /e/, on the other hand, the phoneme boundary shifted inward from the center frequencies.

The results for the vowels /i/, /a/, and /o/ when the synthesizer was excited by a noise are shown in Fig. 4. It was found that the perceptual phoneme boundary shifted in a way similar to that in Fig. 3.

Fig. 5 shows the results when the fundamental frequencies were raised in proportion to the frequency expansion ratios. It was observed that the extent of the shift of the perceptual phoneme boundaries became smaller, and that the center frequencies outside the phoneme boundary in Fig. 3 came back within the boundary.

This result suggests that a certain relation between the formant structure and the fundamental frequency is significant for identifying vowels with various frequency axes correctly.

4. Conclusion

The effects of the frequency expansion, as well as the fundamental frequency raising, on the ranges of F_1 and F_2 in which the frequency expanded vowels were correctly identified as their originals were investigated, using isolated, steady-state synthetic vowel stimuli. The following results were obtained.

(1) The ranges of F_1 and F_2 were generally expanded by the frequency expansion in proportion to the frequency expansion ratios. This result suggests the existence of a framework in which frequency expanded vowels are relatively identified.

(2) The extent of the effect of the frequency expansion varied among vowel classes.

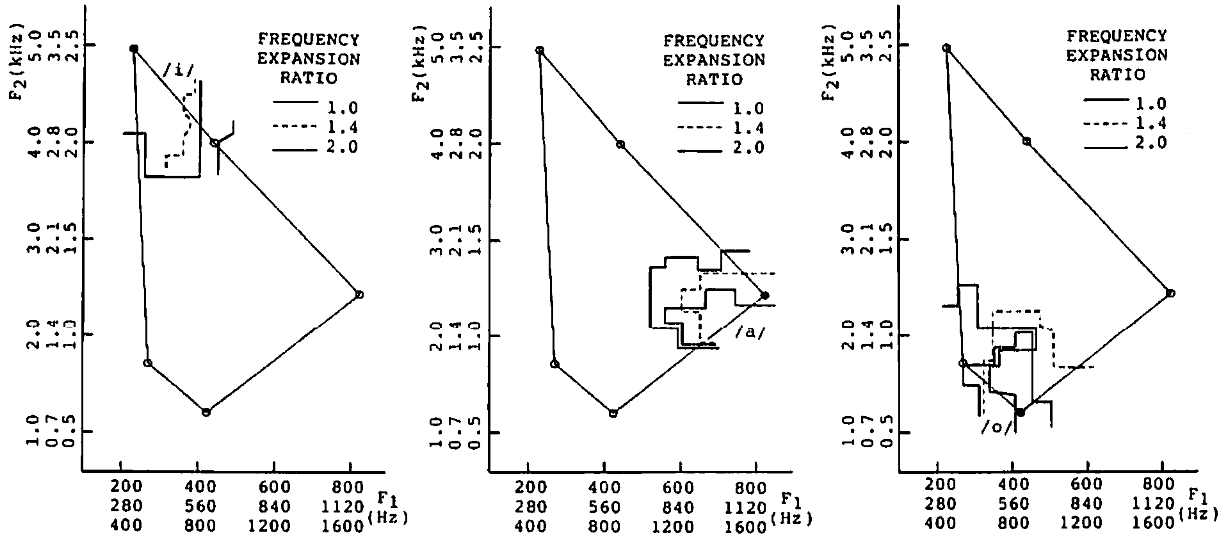


Fig. 4 The shift in the perceptual phoneme boundary when the synthesizer was excited by a noise.

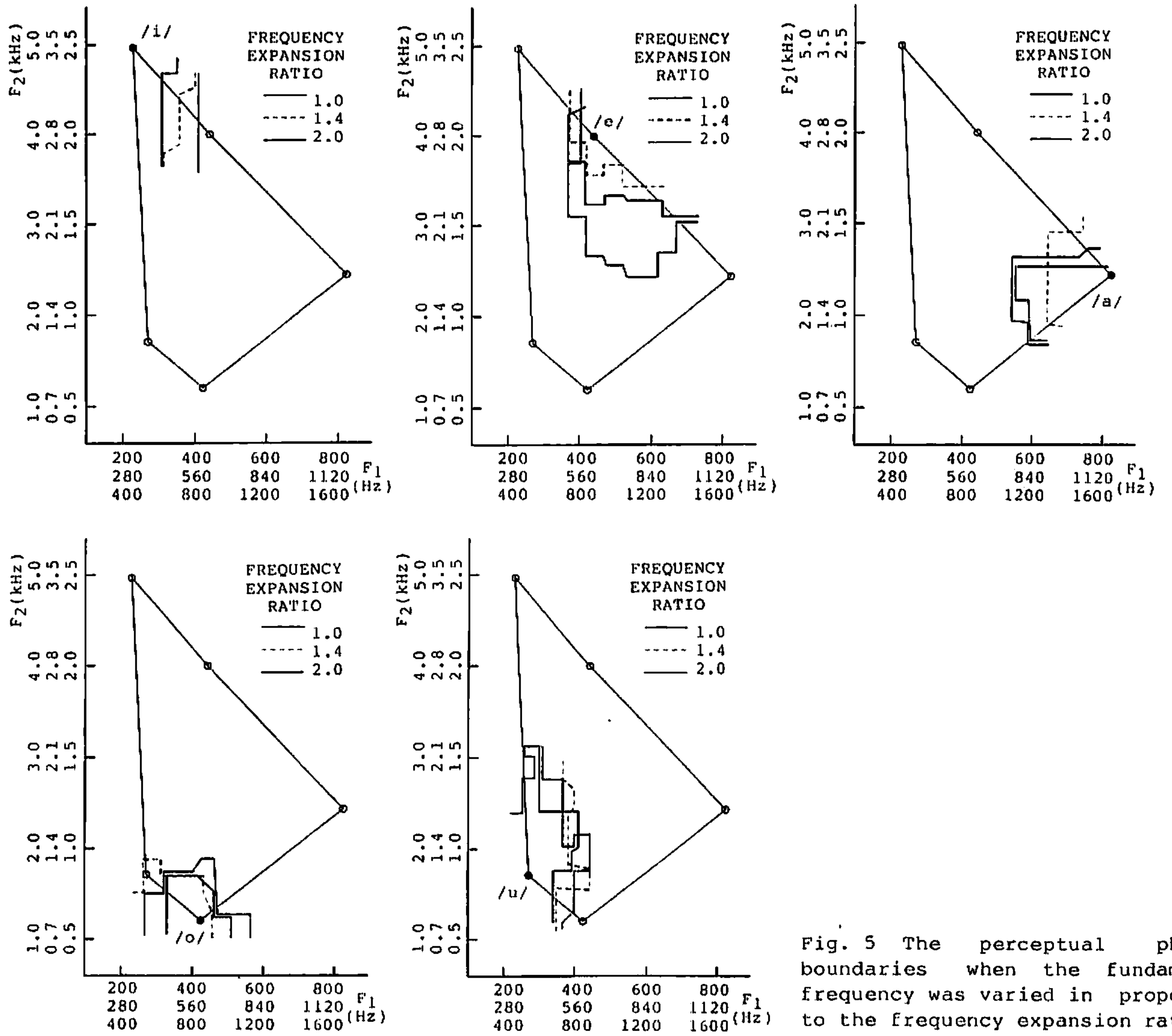


Fig. 5 The perceptual phoneme boundaries when the fundamental frequency was varied in proportion to the frequency expansion ratio.

(3) It was observed that a certain relation between the formant structure and the fundamental frequency was significant for the normalization of the variations in the frequency axis.

References

Fujikaki H. and T. Kawashima (1968); The Roles of Pitch and Higher Formants in the Perception of Vowels. IEEE Trans. Audio Electroacoust., AU-16 (1), 73-77.

Sekimoto S. (1982); Perceptual normalization of frequency scale. Ann. Bull. RILP, 16, 95-101.

Sekimoto S. (1983); Normalization of the speaker difference in vowel perception. Ann. Bull. RILP, 17, 83-96.