ROLES OF PITCH AND HIGHER FORMANTS IN PERCEPTION OF VOWELS

HIROYA FUJISAKI AND TAKAKO KAWASHIMA*

Spectral analysis of the Japanese vowels shows that the five vowels [*a*], [*e*], [*i*], [*o*] and [*u*] of a single speaker can be separated by their first and second formant frequencies (F1 and F2). Considerable amount of overlap is observed, however, between [*u*] and [*e*] as well as between [*o*] and [*a*], when vowels of many speakers are plotted on F1—F2 plane, as in Fig. 1. The two vowels in each pair share approximately the

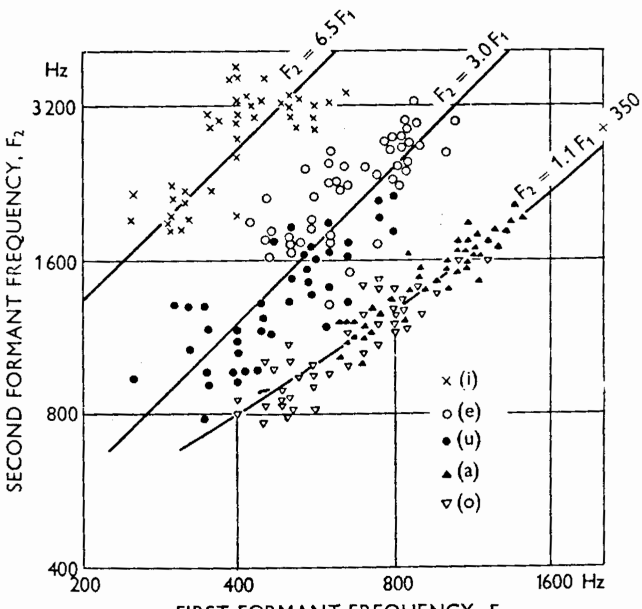

Fig. 1. F_1 — F_2 diagram of the Japanese vowels.

same ratio of F2 to F1, and the overlap can be ascribed mainly to differences in the length of vocal tract. A normalizing process, based presumably on higher formant frequencies (F3, F4 etc.), is expected in the identification of these vowels. It is not

* Department of Electronic Engineering, University of Tokyo.

clear, however, whether concurrent changes of pitch frequency (F0) and formant frequencies, as observed in Fig. 2, are necessary in the normalizing process. In order to estimate the importance of these parameters as perceptual cues, the following method has been adopted using synthetic vowels.

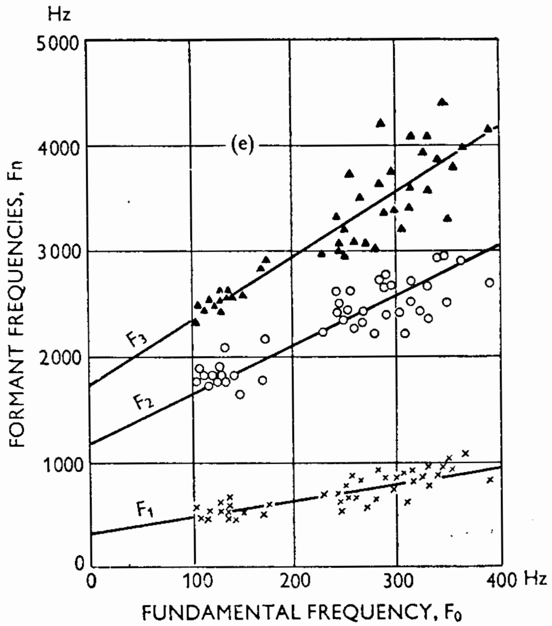

Fig. 2. Fundamental frequency F_0 versus formant frequencies F_1, F_2 and F_3 for [*e*].

Method. The frequency parameters of synthetic vowels generated by the terminal analog synthesizer of Fig. 3 are F0, F1, F2, F3, F4 and Kr4 (higher pole correction), all of which are variable. The spectral envelope of buzz source has no zeros and falls with a slope of —12 dB/oct. from 170 Hz. When F1 and F2 are varied along the direction of overlap, (for example, F2/F1 = 3 for [*u*] and [*e*]), keeping all other parameters constant, a perceptual transition is observed between the two vowels (Fig. 4), and the values of F1 and F2 corresponding to 50 % responses can be defined as the boundary values F1b and F2b.

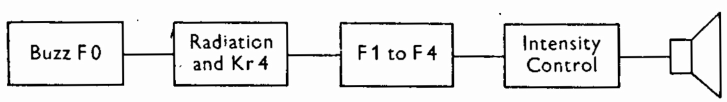

Fig. 3. Block diagram of the vowel synthesizer.

In the following experiments, the effects of i) changes of F3, F4 and Kr4 as a group, ii) change of F0 only, and iii) concurrent changes of F0, F3, F4 and Kr4 were examined. In case iii), an experimental relationship F3 = 6.2 (F0 + 270) Hz was adopted, and F3, F4 and Kr4 were varied always following the pattern for the neutral vowel.

The ranges for F0, F3 and F4 were 140 — 345 Hz, 2.5—4.0 kHz, and 3.5—5.6 kHz, respectively. The duration of the stimuli was 0.8 sec and a rise time constant of 20 msec was given to pitch and intensity for naturalness. For the purpose of comparison, the effects of higher formant changes were also examined for vowel sounds excited by a noise source which had a flat spectrum up to 8 kHz.

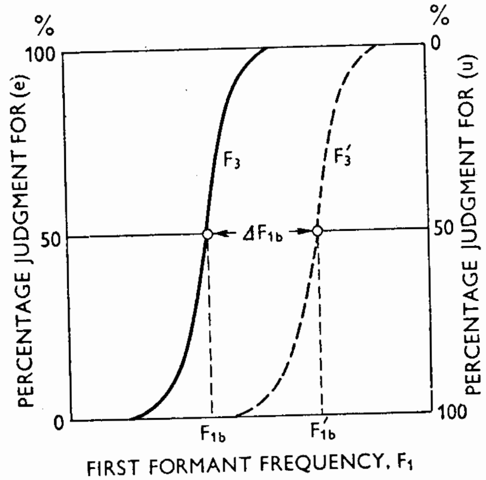

Fig. 4. Perceptual transition between [*u*] and [*e*] and the boundary F_{1b} as affected by a change in F_3.

Results. Table 1 summarizes the results of listening tests performed on 14 subjects (10 adults and 4 children). Since no significant differences were observed due to age or sex, mean values for all subjects are listed. The results show that, for [*u*] — [*e*] identification, the rôles of higher formants are not so important as those of

Table 1. Influence of F0, F3, F4 and Kr4 on the perceptual boundaries between [u]—[e] and [o]—[a].

Source			[*u*]—[*e*]	[*o*]—[*a*]
Buzz	i)	$\frac{\Delta F1b}{F1b}$ $\frac{\Delta F3}{F3}$	0.63	0.13
	ii)	$\frac{\Delta F1b}{F1b}$ $\frac{\Delta F0}{F0}$	0.22	0.29
	iii)	$\frac{\Delta F1b}{F1b}$ $\frac{\Delta F3}{F3}$	1.00	0,93
Noise		$\frac{\Delta F1b}{F1b}$ $\frac{\Delta F3}{F3}$	0.96	0.76

Table 1. Influences of F0, F3, F4 and Kr4 on the perceptual boundaries between [*u*] — [*e*] and [*o*] — [*a*].

F1 and F2, and concurrent changes of pitch and higher formants are necessary to counteract the changes in F1 and F2. For [*o*] — [*a*] identification, the tendency is still greater and the influences of higher formants are even less than that of pitch alone. In noise-excited vowels, on the other hand, the rôles of higher formants are as important as the combined rôles of pitch and higher formants in buzz-excited vowels.

Conclusion. A method has been described to examine the rôles of pitch and higher formants in the perception of vowels. Results of experiments have revealed that, for ordinary buzz-excited vowels, perceptual normalization is not complete unless pitch and higher formants vary concurrently.