ACOUSTICAL LETTER

On the quality of vowels having extraordinary formant frequencies

Tetsuya Watanabe, Tohru Ifukube, Takashi Izumi, and Toshiki Imamura

Research Institute for Electronic Science, Hokkaido University,

Kita-12, Nishi-6, Kita-ku, Sapporo, 060 Japan

(Received 21 December 1992)

Keywords: Extraordinary vowel, Formant, Vowel quality, Bark scale, Synthesized sound
PACS number: 43.71.Es

1. Introduction

It is a well-known fact that some sounds are often perceived as vowels although their formant frequencies are different from those of ordinary vowels. For example, Suzuki *et al.*¹⁾ reported that while the formant frequencies of the vowels produced in a high-pressure, helium-oxygen mixture were more than twice as much as those produced in a normal air, the intelligibility kept 70%. In this case, the question remains as to how the vowel quality is related to the formant frequencies. In order to determine this relationship, a vowel identification experiment was carried out using synthesized sounds.

2. Method

2.1 Sound stimuli

Figure 1 shows the total eighty-six stimuli. Thirty-seven of them ranged from 0.5 to 3.5 times of the formant frequencies and bandwidths of the three Japanese vowels (/i/, /e/ and /a/) uttered by a native Japanese

speaker.²⁾ The rest of the eighty-six stimuli were designed so as to interpolate the above thirty-seven stimuli on the first versus the second formant frequency plane (F_1 - F_2 plane). The stimuli were synthesized using a Klatt-type cascade three-formant synthesizer³⁾ with a sampling rate of 25 kHz. The duration of stimuli was 600 ms including a 100 ms on- and off-glide. The fundamental frequency was fixed at 123 Hz.

2.2 System

Figure 2 shows a block diagram of the vowel identification experiment. Sound stimulation patterns memorized in the personal computer were converted into analog signals. Then the signals were filtered through a low-pass filter and amplified. Subjects were asked to listen to the stimuli binaurally through headphones.

2.3 Subjects

Subjects were five male and five female native Japanese aged 23 to 27 years. All of them have normal hearing.

2.4 Procedure

Prior to the experiment subjects adjusted the volume so that they did not have any difficulty in listening. Subjects listened to five sets of stimuli. Each set consisted of eighty-six sound stimuli which were randomly arranged. They were asked to identify the stimuli as one of the five Japanese vowels or as a non-vowel sound.

Fig. 1 Positions of eighty-six sound stimuli on the F_1 - F_2 plane on a linear scale. Three letters in the plane, i, e and a, represent the original data. The curved line distinguishes the normal (see Ref. 4) from extraordinary vowels.

Fig. 2 Block diagram of the vowels identification experiment.

3. Results and discussion

Figure 3 shows the distribution of the stimuli which were identified as the five Japanese vowels on the F_1 - F_2 plane on a linear scale. Large bold letters represent the stimuli which were identified as a certain vowel at a rate of more than 80%, and small plain letters at a rate of from 61 to 80%. The lines of the perception regions enclose all of the same letters.

In Fig. 3 two features are found:

(1) All of the sound stimuli whose formant frequencies are higher than the ordinary spoken vowels were identified as /i/, /e/ or /a/.

Fig. 3 Distribution of the stimuli which were identified as the five Japanese vowels on the F_1 - F_2 plane on a linear scale.

Fig. 4 Distribution of the stimuli which were identified as the five Japanese vowels on the F_1 - F_2 plane on the Bark scale.

(2) The perception regions of /i/, /e/ and /a/ extend upward towards the left on the F_1 - F_2 plane. In other words, the ratio of the second to the first formant frequency of the stimuli which have higher formant frequencies is larger than that of normal vowels.

From the second feature, it may be deduced that a certain amount of frequency change in a higher frequency region (over 3,000 Hz as seen in Fig. 3) has less effect on vowel identification than the same amount in a lower frequency region. This hypothesis can be supported by Zwicker's experimental result that the critical-band width becomes wider as the frequency increases.⁵⁾ It is therefore reasonable to use the Bark scale that corresponds to the critical-band width instead of a linear scale when discussing vowel identification. Figure 4 represents the results on the Bark scale.

In Fig. 4 the stimuli that were identified as /i/, /e/ or /a/ are not mixed but clustering respectively. This means that the vowel quality can be determined by the position of the stimuli on the F_1 - F_2 plane even if both frequencies are much higher than those of ordinary spoken vowels.

4. Summary

In order to understand how higher formant frequencies affect the vowel quality, a vowel identification test was carried out using synthesized sound stimuli. The results summarized are:

(1) The sound stimuli whose formant frequencies are higher than the ordinary spoken vowels were all identified as /i/, /e/ or /a/.

(2) From their distribution on the F_1 - F_2 plane on the Bark scale, it is concluded that the vowel quality is determined by the first and the second formant frequencies even if both frequencies are much higher than those of ordinary spoken vowels.

Before going into further discussion on this matter, another experiment needs to be conducted using the stimuli that scatter uniformly on the Bark scale.

References

- 1) J. Suzuki, M. Nakatsui, T. Takasugi, and R. Tanaka, "Final report on helium speech investigations," *Rev. Radio Res. Labs.* **24** (129) 394 (1978), (in Japanese).
- 2) T. Hirahara and H. Kato, "On the role of fundamental frequency in steady vowel perception," *IEICE Tech. Rep.* 88-126, 33-40 (1988).
- 3) D. H. Klatt, "Software for a cascade/parallel formant synthesizer," *J. Acoust. Soc. Am.* **67**, 971-995 (1980).
- 4) H. Kasuya, H. Suzuki, and K. K. Kido, "Changes in pitch and first three formant frequencies of five Japanese vowels with age and sex of speakers," *J. Acoust. Soc. Jpn. (J)* **24**, 355-364 (1968) (in Japanese).
- 5) E. Zwicker, G. Flottorp, and S. S. Stevens, "Critical bandwidth in loudness summation," *J. Acoust. Soc. Am.* **29**, 548-577 (1957).