Ann. Bull. RILP (1982) No. 16, 95-101

PERCEPTUAL NORMALIZATION OF FREQUENCY SCALE

Sotaro Sekimoto

1. Introduction

It is well known that vowel quality differences can be described in terms of the lowest two formant frequencies, F_1 and F_2. However, considerable variation is found in the position of vowels on the F_1-F_2 plane depending upon the speaker. The position of some vowels on the F_1-F_2 plane uttered by one speaker sometimes overlaps with that of different vowels uttered by other speaker. These overlaps are found especially between male and female voices. It is supposed that some normalization mechanism exists in the perceptual process in order to compensate for these formant variations in identifying vowels.

In a study by the present authors on the frequency compression technique for hearing aids, it was observed that similar overlaps in formant position occur among the different compression rates in frequency compressed speech[1]. It is, therefore, suggested that a similar normalization mechanism could play a role in removing these overlaps in the perceptual process of recognizing a target vowel.

In this study, three experiments were carried out to clarify the normalization mechanism. In the experiments, the frequency axis was expanded or compressed in order to simulate the positional variation of vowels on the F_1-F_2 plane.

2. Method

2.1 Conditions

The experimental conditions employed in Experiment I, II, and III are shown in Table 1.

(1) Frequency expansion or compression ratio

The frequency expansion or compression ratio was defined as the percent ratio of the maximum frequency range of frequency expanded or compressed speech to that of noncompressed speech.

(2) Fundamental Frequency

(i) Original F_0: In this case, fundamental frequency was maintained at the same value as the original.
(ii) Proportional F_0: In this case, fundamental frequency was raised or lowered in proportion to the frequency expansion or compression ratio, respectively.

2.2 Frequency Expansion or Compression Method

Frequency expansion or compression was performed, using a PARCOR speech analysis-synthesis system, by altering the sampling frequency. The analysis was made every 5 msec on 30 msec of Hamming windowed speech. The order of the

Table 1 *Experimental Conditions*

	Frequency Expansion or Compression ratio	Fundamental Frequency
EXP. I	50, 60, 80, 100, 140, 180, 220 (both with male and female voice), and 260 (only with male voice)	ORIGINAL and PROPORTIONAL
EXP. II	180 and 220 (female voice) 220 and 260 (male voice)	PROPORTIONAL
Exp. III	50, 60, 80, 100, 140, 160, 180, 200, 220, and 240	ORIGINAL and PROPORTIONAL

PARCOR analysis was 12. The pitch period was determined by detecting the autocorrelation peak of the residual wave with the spectral deformation caused by the vocal tract resonance removed.

2.3 Procedure

The speech materials were five steady state Japanese vowels uttered by two female speakers and one male. The speech samples processed by the method described above were subjected to an intelligibility test. The tests were carried out in a soundproof room for every combination of the experimental conditions and speakers. The speech stimuli were presented in random order to one ear by a headphone at a level of 10 dB OTR (Ortho-telephonic response). Three male and four female subjects were requested to identify each stimulus as one of the five Japanese vowels.

3. Experiment I

In Experiment I, the vowel intelligibilities were measured for the different frequency expansion or compression ratios with the fundamental frequency condition. The results for one female speaker are, for example, shown in Fig. 1, where the vowel identification score is plotted against the frequency expansion or compression ratio. The filled circles show the result with the original F_0 and the open circles show the result with the proportional F_0. The frequency expansion or compression ratio with which the vowels were identified correctly was found to be between 80 and 180% and between 50 and 220% for the original F_0 and the proportional F_0 conditions, respectively.

Fig. 2 shows the results for the male speaker. The range of the expansion or compression ratio with which the vowels were identified correctly was 80–260% for the original F_0, and 80–260% for the proportional F_0. The upper limit of the

frequency expansion or compression ratio was found to be greater for male speech than for female speech.

Fig. 3 shows the vowel positions on the F_1-F_2 plane for the same female speaker as shown in Fig. 1. Vowels with the same frequency expansion or compression ratio are connected with a line. In the figure, overlaps are, for example, observed in the region of /u/ and /e/, and /a/ and /o/ when different frequency expansion or compression ratios are applied. These results imply that some frequency normalization function exists in the perceptual process, and that the value of the fundamental frequency relative to the formant frequencies contributes to the normalization.

4. Experiment II

Experiment II was carried out to examine whether identification of frequency expanded or compressed vowels is made based on the relative formant position of each vowel which has the same formant expansion or compression ratio, or based on the absolute formant position. The intelligibility scores for the group of frequency expanded vowels were compared with the scores for both frequency expanded and non-expanded vowels.

The highest two ratios of the frequency expansion or compression, i.e., 180% and 220% for the female voice, and 220% and 260% for the male voice, were used. The fundamental frequency was raised in proportion to the frequency expansion ratio.

Fig. 4 shows the relationship between vowel intelligibility and the frequency expansion or compression ratio. The intelligibilities for the frequency expanded vowel group, shown by solid lines, were found to be 100%. On the other hand, the intelligibilities for both frequency expanded and non-expanded vowels, shown by broken lines, were rather low. The more the frequency was expanded, the greater the intelligibility was reduced. The intelligibility of the non-expanded vowels which were mixed with the expanded vowels was 100%, namely their intelligibility scores were not affected by the addition of the expanded vowels. These results suggest that frequency expanded vowels are identified in a absolute manner when the frequency scaling conditions are natural, i.e. when the frequency expansion ratio is fixed at around 100%. When the frequency expansion or compression ratio becomes greater, however, frequency expanded vowels seem to be identified by means of some relative frame. In other words, there seem to exist framework effects concerning the frequency expansion ratio.

5. Experiment III

In Experiment III, the effects of the fundamental frequency were investigated more closely. Since Experiment II showed a framework effect for the expansion or compression ratio, the vowel identification measurements were conducted separately for each vowel in order to exclude these frame effects. The step size of the ratio for the frequency expansion was changed to half of those used in Experiment I, whereas the same step size of the ratio was used for the frequency compression. The ratios are shown in Table 1. The fundamental frequency conditions were the

same as those used in Experiment I.

Fig. 5 shows the effect of the fundamental frequency on the identification of the vowel /o/, where identification rates are plotted against the freqeuncy expansion or compression ratio. The results for the original F_0 are indicated by filled circles, and for the proportional F_0 by open circles. For the original F_0, the vowel changed as the frequency expansion or compression ratio varied. On the other hand, the vowel /o/ was identified correctly for all of the frequency expansion or compression ratios when the fundamental frequency was raised or lowered proportionally. The vowel /o/ tended to be confused with one of the non-expanded or non-compressed vowels which had almost the same F_1 when the fundamental frequency was kept unchanged. This tendency was, however, not clearly present for the other vowels.

6. Comments

As cues for the frequency normalization for expanded or compressed speech, the presence of a framework effect within the same frequency expansion or compression ratio and an effect for fundamental frequency were found. As other factors might exist, further study is now in progress.

Acknowledgment

This study was supported in part by a Grant-in-Aid for Scientific Research (No. 57780024) from the Japanese Ministry of Education, Science and Culture.

References

1. Sekimoto, S. (1982); Effects of Formant Peak Emphasis on Vowel Intelligibility in Frequency Compressed Speech, Ann. Bull. RILP. 16. 89-94.
2. Sekimoto, S., S. Kiritani and S. Saito (1980); Intelligibility of frequency compressed speech in low-pass filtered condition, Ann. Bull. RILP, 14, 181-193.
3. Sekimoto, S., S. Kiritani and S. Saito (1980); Perceptual phoneme confusion of frequency compressed speech, Ann. Bull. RILP, 14, 195-200.
4. Sekimoto, S., S. Kiritani and S. Saito (1979); Frequency compression technique of speech using linear prediction analysis-synthesis scheme, Ann. Bull. RILP, 13, 133-136.

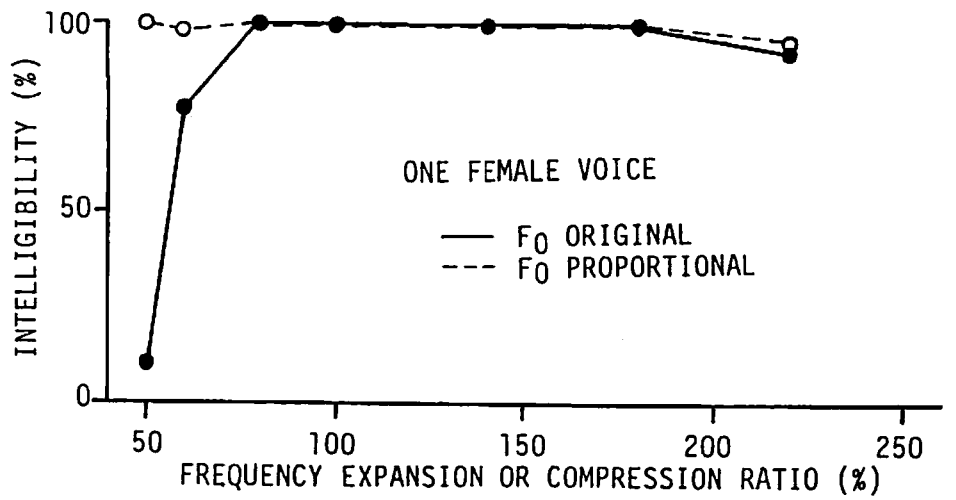

Fig. 1 *Relationship between vowel intelligibility and the frequency expansion or compression ratios. The results for one female voice are plotted with the two fundamental conditions: original* F_0 *and proportional* F_0.

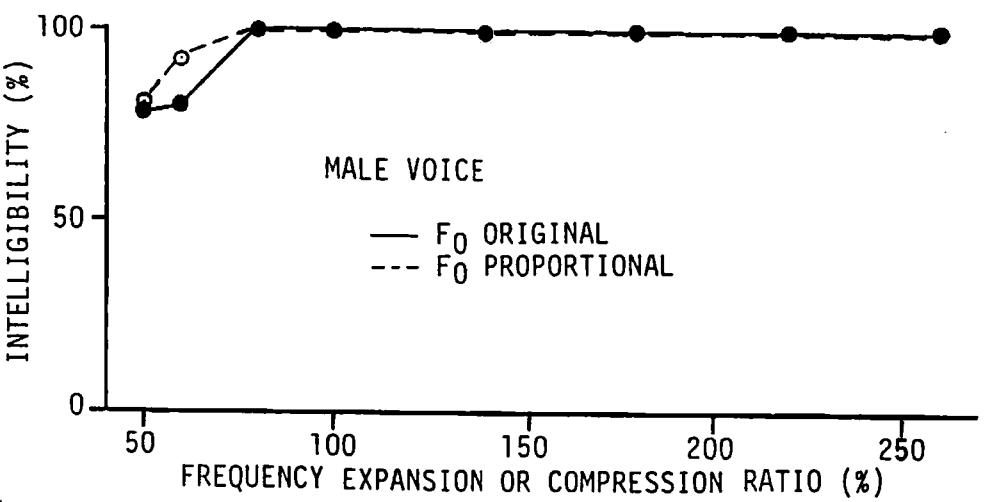

Fig. 2 *Relationships between vowel intelligibility and the frequency expansion or compression ratios. The results for the male voice are plotted with the two fundamental conditions: original* F_0 *and proportional* F_0.

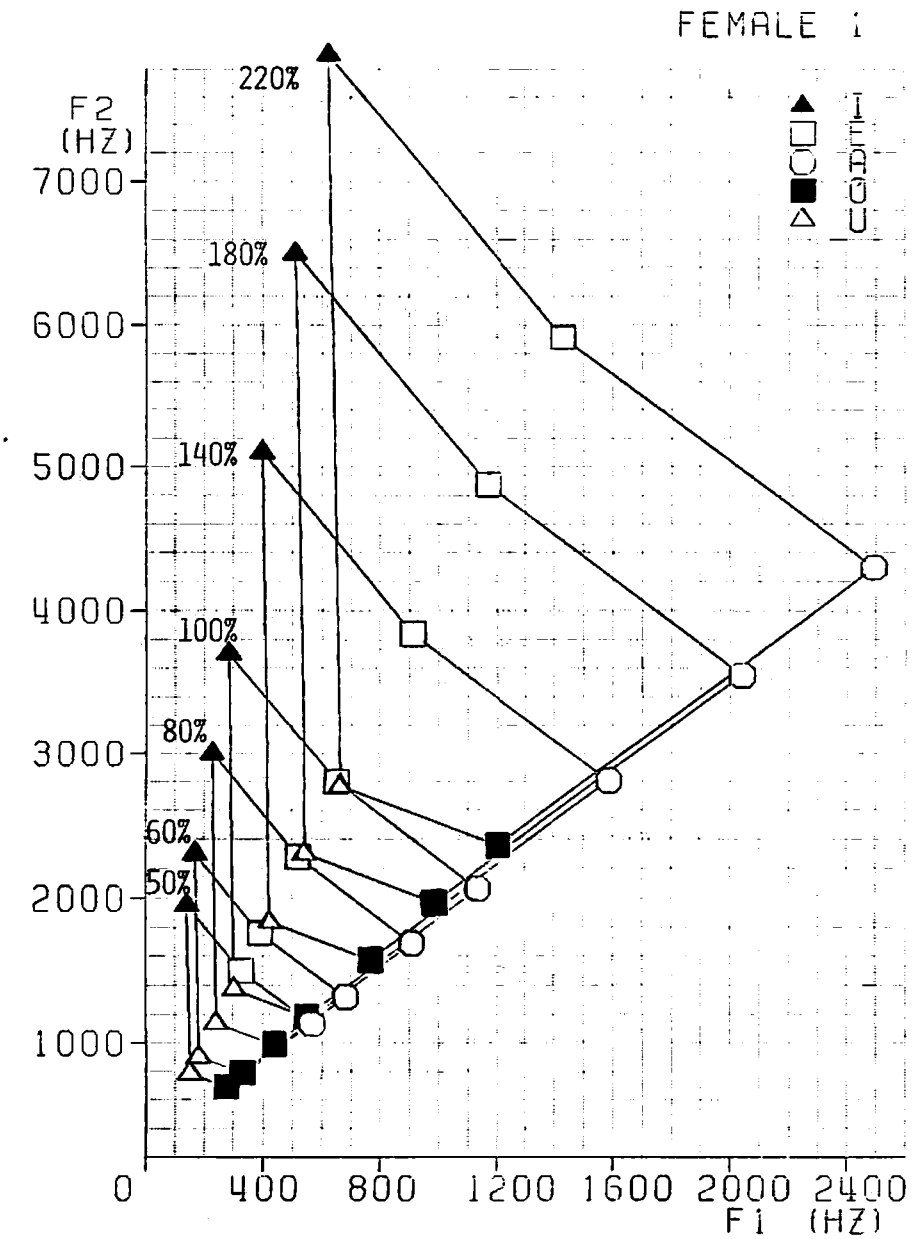

Fig. 3 *F_1-F_2 representation of the five Japanese vowels uttered by one female speaker. The pentagonal shaped lines encompass the same frequency expansion or compression ratio.*

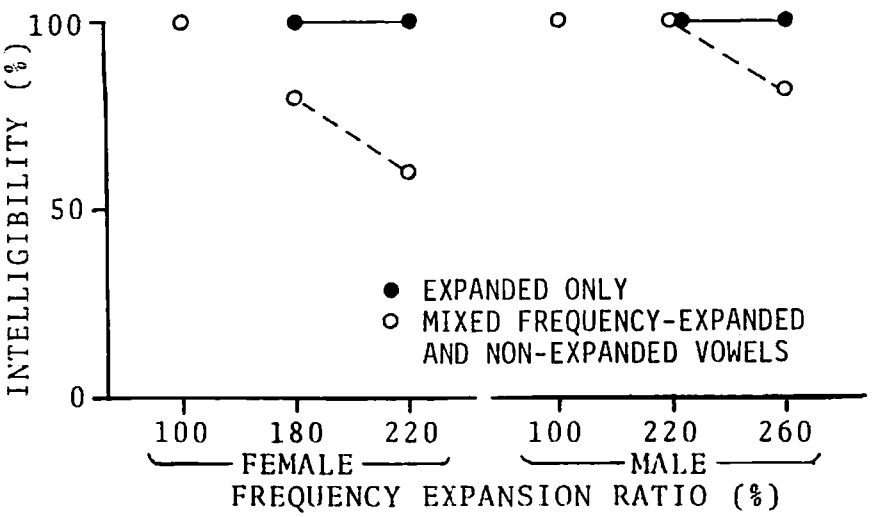

Fig. 4 *Vowel intelligibility for frequency expanded speech. Here, a comparison is made between when frequency expanded vowels were presented alone and when non-compressed vowels were added.*

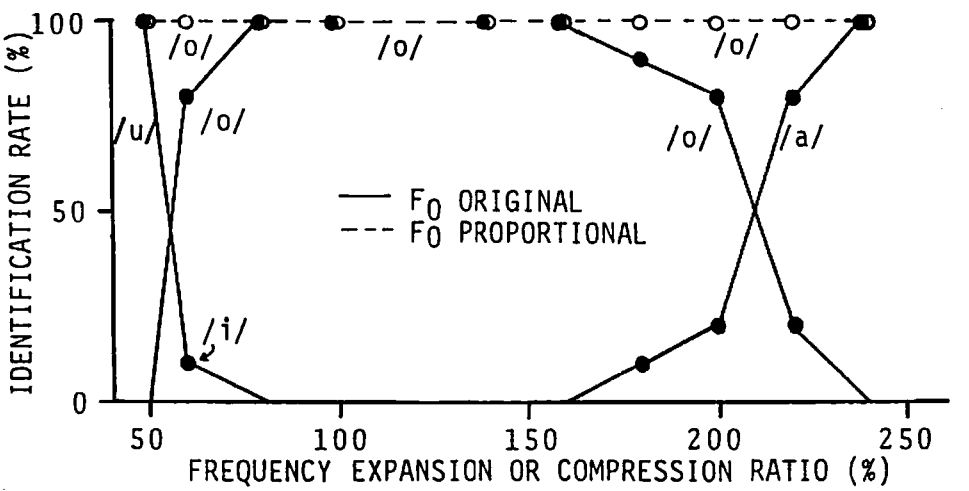

Fig. 5 *Effect of the fundamental frequency lowering on the identification of the vowel /o/.*