Ann. Bull. RILP, No. 14, 195-200 (1980)

PERCEPTUAL PHONEME CONFUSION OF FREQUENCY-COMPRESSED SPEECH

Sotaro Sekimoto, Shigeru Kiritani and Shuzo Saito

1. Introduction

A new frequency compression technique based on the PARCOR speech analysis-synthesis method has been proposed by the present authors[1], and intelligibility tests of frequency-compressed speech have been conducted using normal hearing subjects with simulated hearing loss. The effects of frequency compression ratios and the values of fundamental frequency on the articulation scores are reported elsewhere[2]. In the present paper, the characteristics of phoneme confusion in the frequency-compressed speech are discussed.

2. Procedure

As the details of the experimental procedure are described elsewhere[2], the experimental conditions of the intelligibility tests are briefly summarized below.

The frequency-compressed speech of 100 Japanese monosyllables uttered by three speakers, one male and two females, was subjected to intelligibility tests. The frequency compression ratios, defined as percent ratios between the bandwidth of the compressed speech and that of original speech, were: 100%, 80%, 60%, 50% and 40%. The fundamental frequencies were: (i) kept at the same value as the original one, and (ii) lowered at the same ratio as that of the frequency compression ratios, respectively. Only the data for the latter condition will be presented in this paper. The cutoff frequencies of the lowpass filter which simulates the frequency characteristics of hearing loss were: (i) "Through," (ii) 1.5kHz, (iii) 1.0kHz and (iv) 0.7kHz, where "Through" means that the lowpass filter was not passed. So the bandwidths of the speech signal were 5.0kHz, 4.0kHz, 3.0kHz, 2.5kHz and 2.0kHz, respectively, for the five different frequency compression ratios.

The synthesized speech sounds were presented in random order to two subjects, one male and one female (different from the speakers), through headphones, in a soundproof room. The subjects were asked to identify the synthetic monosyllables as one of 100 Japanese monosyllables.

3. Confusion of Vowel Sounds of Frequency-Compressed Speech

The articulation scores of individual vowels are shown in Figures 1–3 for two conditions of lowpass filtering: "Through" and 0.7kHz.

For the female voices, it is observed in Fig. 1 and Fig. 2 that in the case of the filtering condition "Through," the articulation score of /i/ becomes lower when the frequency compression ratio is below 60%. Analysis

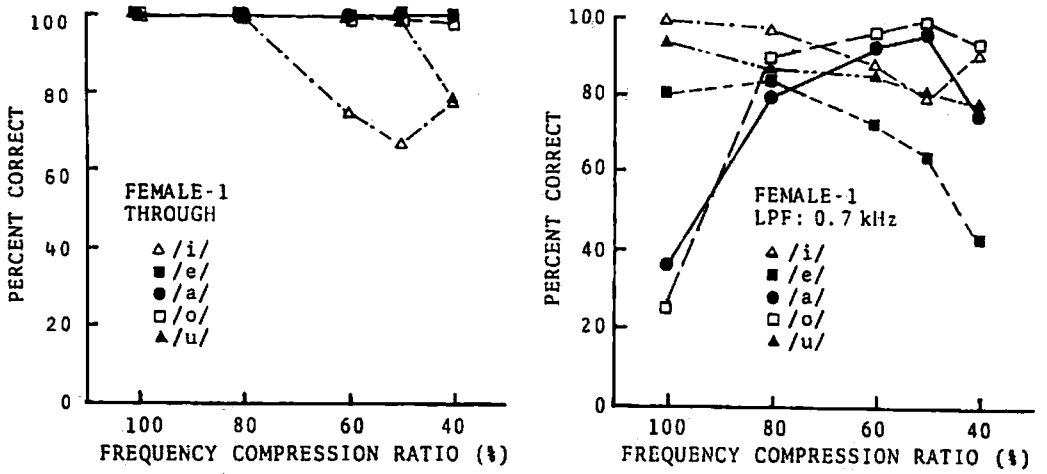

Fig. 1 Relationships between frequency compression ratios and the articulation scores for each vowel obtained for female-1 voice.

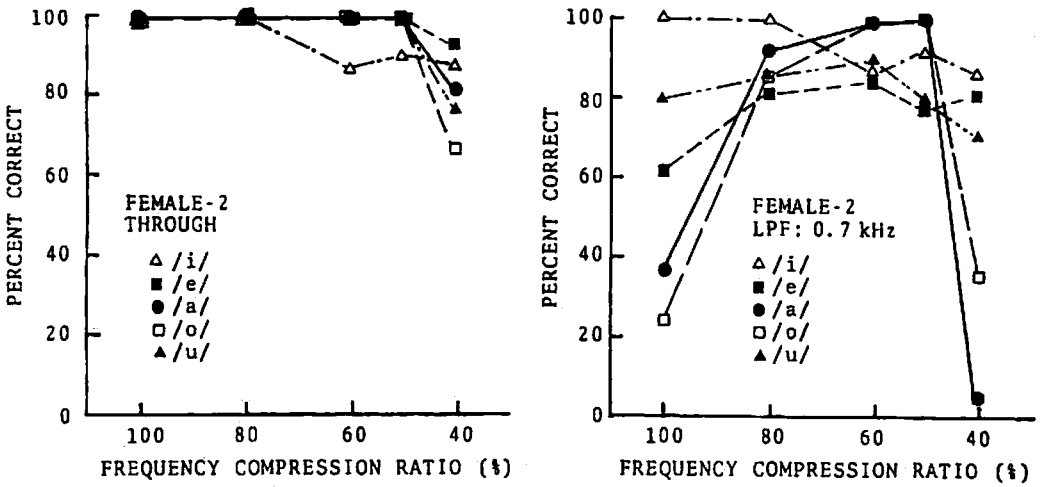

Fig. 2 Relationships between frequency compression ratios and the articulation scores for each vowel obtained for female-2 voice.

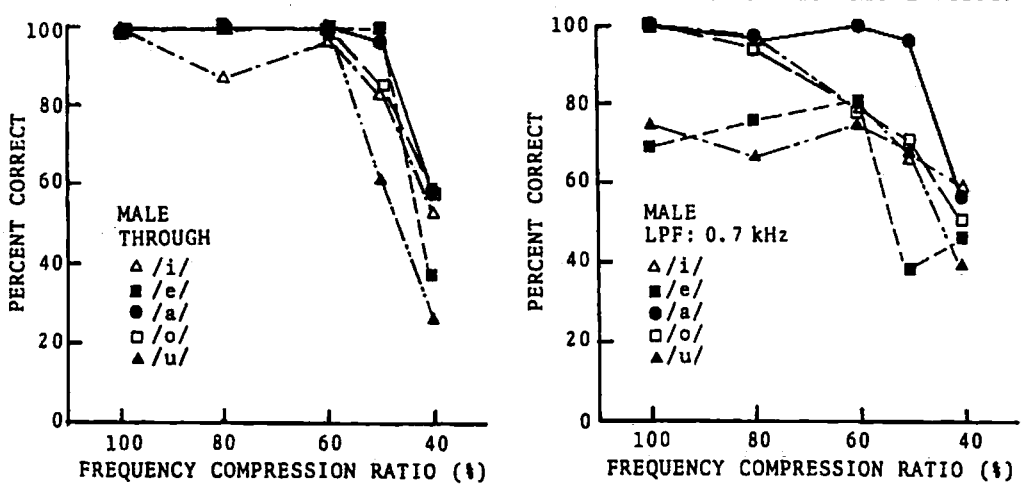

Fig. 3 Relationships between frequency compression ratios and the articulation scores for each vowel obtained for male voice.

of the confusion matrix shows that /i/ is confused with /u/. Identification of the other vowels is almost perfect except when the frequency compression ratio is 40%. In the case of the filtering condition at 0.7kHz, the articulation scores for /a/ and /o/ are 36% and 26%, respectively, for the original speech (at 100% frequency compression ratio). The error involved with these vowels is mainly confusion with /e/. These confusions are decreased and the articulation scores are improved up to 96% by frequency compression with a ratio of 50%. On the other hand, the articulation scores for the other vowels, /i/, /u/ and /e/, are lowered by this frequency compression.

For the male voices, improvement in the articulation score after frequency compression is not observed. In the case of the filtering condition "Through," /i/ is sometimes confused with /u/ or /e/. In the case of the filtering condition at 0.7kHz, /u/ is confused mainly with /i/, and /e/ is confused mainly with /o/.

4. Confusion Characteristics of Consonant Sounds

As the confusion matrices of consonants showed very complicated patterns, analysis was made of the patterns of confusion between the various groups of consonants. The consonants are classified according to acoustic characteristics, voiced/voiceless feature and a feature related to spectral patterns. The classifications of the consonants used here are shown in Table 1.

Table 1. Classification of Consonants

SPECTRUM FEATURE \ VOICED/VOICELESS	VOICED	VOICELESS
GROUP 1	b bj w m mj	p pj
GROUP 2	d dz dʒ n nj j	t tj ts s sj
GROUP 3	g gj r rj	k kj h hj

This is a classification system proposed by Saito for the analysis of confusion matrix among the Japanese consonants.

4.1 Confusions between Voiced and Voiceless Consonants

The rate of correct identification of the voiced/voiceless feature in the lowpass filtering conditions "Through" and 1.5kHz are shown in Figure 4. The abscissa indicates the frequency compression ratio, and the ordinate indicates the average articulation score for the two subjects. It can been seen in these figures that:

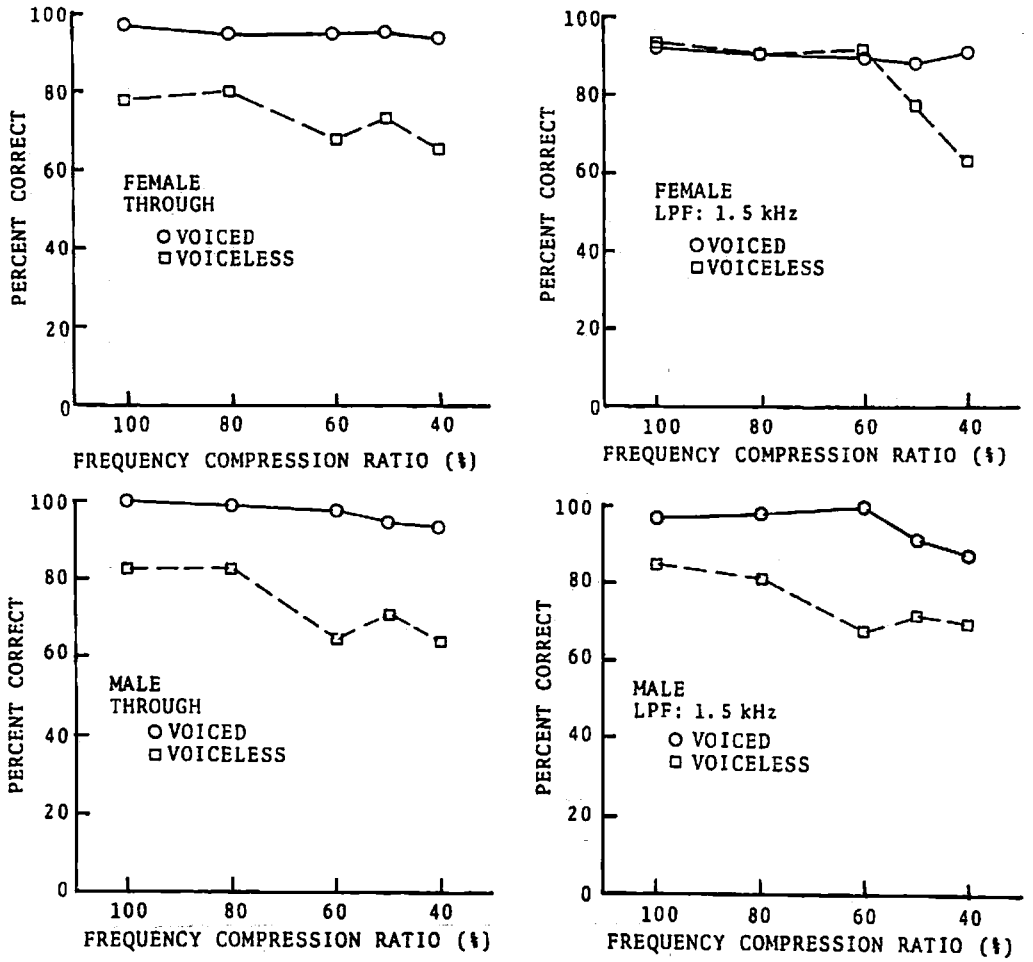

Fig. 4 Relationships between frequency compression ratios and the correct identification rates for voiced/voiceless feature.

(1) The perceptual error of the voiced/voiceless feature is very low, and most of the errors are 'voiced' response for the voiceless consonants.
(2) The frequency compression does not affect the identification of the voiced/voiceless feature.

4.2 Confusions Concerning the Spectrum Feature

Figure 5 shows the articulation score for the spectrum feature in the three lowpass filtering conditions "Through," 1.5kHz and 0.7kHz. The abscissa indicates the frequency compression ratio, and the ordinate indicates the rate of identification in which the consonant belonging to group i was judged as one of the consonants belonging to the same group. The averages of the correct identification rates for the two subjects are shown. The following results are observed from these figures.

(1) In the case of the lowpass condition "Through," all articulation scores are lowered by frequency compression.

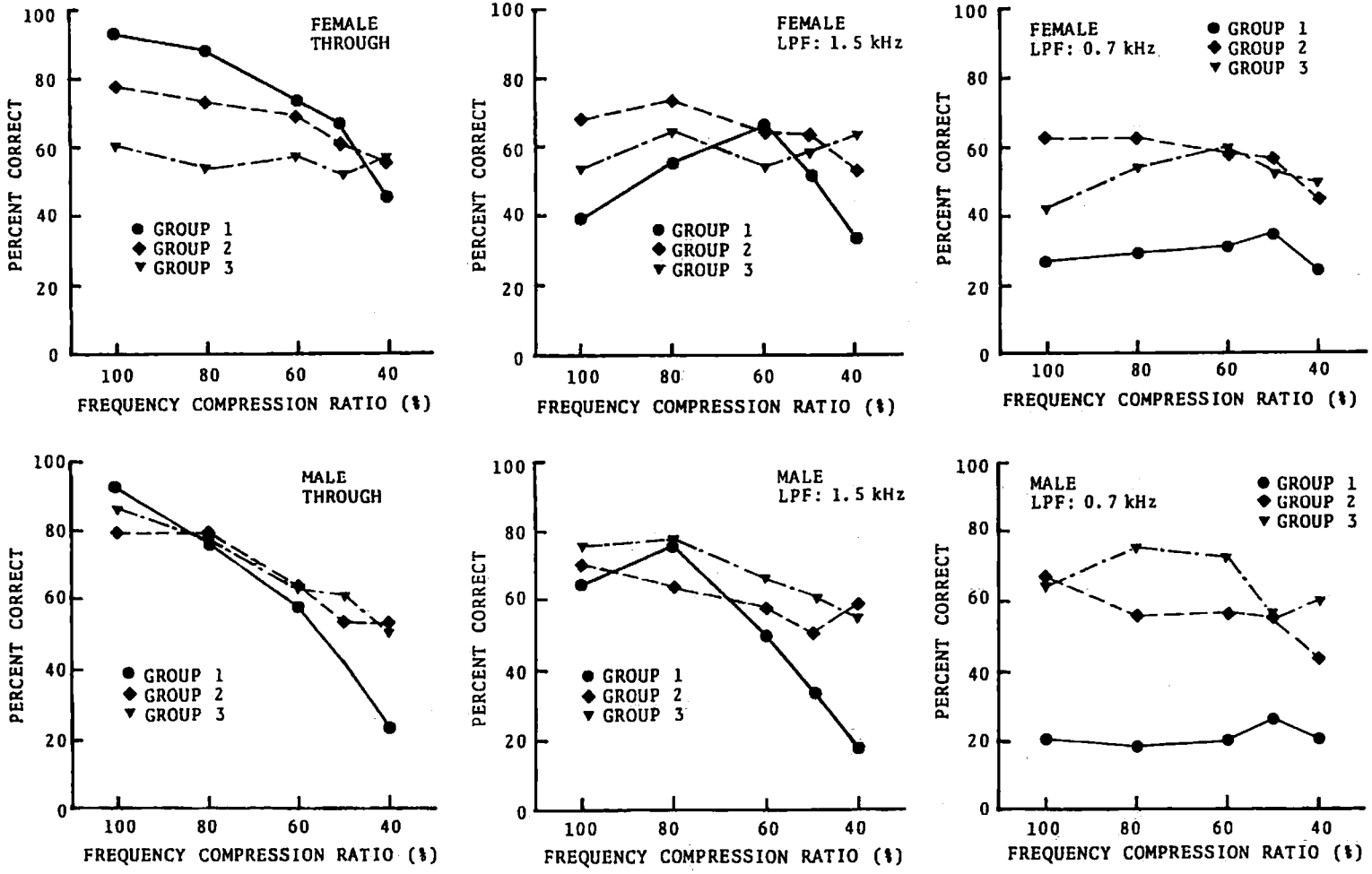

Fig. 5 Relationships between frequency compression ratios and the correct identification rates for spectrum feature.

(2) In the case of the 1.5kHz lowpass condition, the articulation score for "Group 1" is improved by frequency compression.
(3) In the case of the 0.7kHz lowpass condition, the articulation score for "Group 3" is improved by frequency compression.

5. Conclusion

The patterns of confusion among vowels and among the class of consonants were analyzed, and the following tendencies were observed.
(1) In the case of the female vowel, lowering of the articulation scores for /a/ and /o/ under the lowpass filtering condition is improved by frequency compression.
(2) In the case of male vowels, on the other hand, no improvement is observed with frequency compression.
(3) Identification of the voiced consonants is almost perfect, but voiceless consonants are sometimes judged as voiced. The identification of voiced/voiceless feature is not affected by frequency compression.
(4) The articulation scores for the spectrum feature under the lowpass filtering condition are improved by frequency compression.

Acknowledgment

This work was supported in part by a Grant-in-Aid for Scientific Research (No. 579013, 540003) from the Japanese Ministry of Education and a Grant-in-Aid for Research on Computed Ear from the Japanese Ministry of Science and Technology.

References

1) Sekimoto, S., S. Kiritani and S. Saito (1979): "Frequency compression techniques of speech using linear prediction analysis-synthesis scheme," Ann. Bull. RILP, 13, pp. 133-136.
2) Sekimoto, S., S. Kiritani and S. Saito (1980): "Intelligibility of frequency compressed speech in lowpass filtered condition," Ann. Bull. RILP, 14, 181-193