Ann. Bull. RILP, No. 14, 181-193 (1980)

INTELLIGIBILITY OF FREQUENCY COMPRESSED SPEECH IN LOWPASS FILTERED CONDITION

Sotaro Sekimoto, Shigeru Kiritani and Shuzo Saito

1. Introduction

The high-frequency components of speech sounds carry more intelligibility than do the low-frequency ones, although the former carry less speech power than the latter. Some deaf persons who have high-frequency response loss of hearing show, in general, a greater intelligibility loss than those with a low-frequency response loss. To prevent such an intelligibility loss in hearing, several techniques of frequency compression for transposing high-frequency components of speech into the usable lower-frequency range have previously been studied in order to provide intelligible speech for this type of deaf person.(1) - (4)

It seems, however, that additional distortions in the speech spectrum accompanied these techniques, and the intelligibility of the output speech was thus rather unsatisfactory. In this study, the PARCOR speech analysis and synthesis method was applied for frequency compression of speech signals(5), and the intelligibility of this frequency compressed speech was measured using subjects with normal hearing acuity under simulated deafness conditions.

2. Frequency Compression Technique

The speech signal was passed through an ordinary PARCOR analyzer to extract the PARCOR coefficients, the pitch period, the speech power, and the voiced and unvoiced excitation ratio. These values of the feature parameters of speech were modified and fed to a PARCOR synthesizer in order to tranpose the poles of the frequency spectrum into the lower frequency region.

The simplest process for modification of frequency compression was achieved by reduction of the sampling frequency of the PARCOR synthesizer, while using the original values of the feature parameters derived from PARCOR analysis. For example, the sampling frequency of 10kHz in the PARCOR analyzer was tranformed to 8kHz in the synthesizer, after which the frequency spectrum envelope was compressed linearly at a frequency compression rate of 80%. According to the reduction of the sampling frequency in the synthesizer, the frequency spectrum envelopes were compressed as shown schematically in Fig. 1.

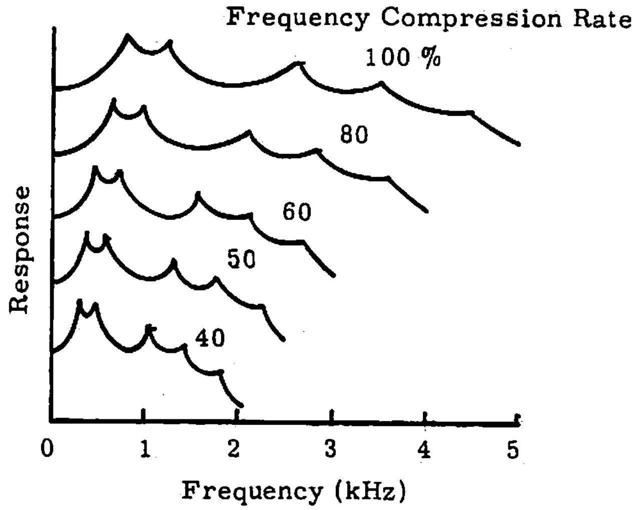

Fig. 1. Frequency spectrum envelopes of speech with various frequency compression rates.

3. Experimental Procedures

3.1 Experimental equipment

The speech signal used for the intelligibility measurement was uttered into a condenser microphone which was placed in front of the talker's lips at a distance of 15cm. The microphone output was amplified and passed through a lowpass filter of 5kHz cutoff frequency. It was then sampled at 10kHz and the amplitude was digitized into 10 bits.

The digitized speech signal was then fed to a PDP-11/34 electronic computer, in which the PARCOR analysis and synthesis system was installed in the software program. The PARCOR coefficients of 12 order were analyzed from the Hamming windowed speech signal of 20 millisecond frame intervals. The analysis was repeated for every 5 millisecond frame period.

The fundamental frequency was derived from the detection of the peak in the autocorrelation of the residual wave, which was derived from the difference between the original and the predicted signals based on the PARCOR coefficients. The residual wave was also used for the determination of the speech power and the voiced and unvoiced excitation ratio, calculating the variance and the interval of the two peaks of the autocorrelation coefficient, respectively.

These feature parameters of the PARCOR analysis were partly modified and then fed to the PARCOR synthesizer, which was installed in the computer. In the synthesizer, the frequency-compressed speech was generated, reducing the sampling frequency and also the number of speech samples per frame interval as described in Sec. 2.

The PARCOR synthesizer had two excitation sources: a pulse generator and a random noise generator. The output levels of both generators were adjusted under the control of the voiced and unvoiced excitation ratio and the speech power signals, and then fed to the cascade-connected digital filters, each being composed of two multipliers, two adders and one delay unit, in which multipliers were controlled by the PARCOR coefficients. The number of the cascade-connected digital filters was the same as the order of the PARCOR coefficients analyzed.

The output of the digital filters was converted to analog speech wave and passed through a lowpass filter, whose cutoff frequency would normally be the same as the input lowpass filter, that is, 5kHz. But, as described in the succeeding section, the cutoff frequency was varied in accordance with the frequency compression rate.

The synthesized output was fed to a second lowpass filter to simulate the high-frequency response loss of a deaf person. It was then sent to the receiver amplifier containing the real ear response equalizer; and finally presented to the listeners monaurally through high-quality earphones. A block diagram of the entire experimental equipment is shown in Fig. 2

3.2 Intelligibility measurement

3.2.1 Frequency compression rate

The frequency compression rates used for the intelligibility test were 100, 80, 60, 50 and 40%, which are executed by the reduction of the sampling frequency in the synthesizer as described in Sec. 2. The highest frequency component of input speech, that is, 5kHz, was reduced at 5, 4, 3, 2.5 and 2kHz, respectively.

3.2.2 Fundamental frequency of the PARCOR synthesizer

For the intelligibility test of the frequency-compressed speech, two kinds of fundamental frequencies of frequency-compressed speech were used: one being the same as the input speech signal and the other being reduced corresponding to the frequency compression rate. In the latter, the fundamental frequency was mostly reduced at the same rate as the frequency compression, but in a few intelligibility measurements, the reduction rate of fundamental frequency was varied independent of the frequency compression rate.

3.2.3 Lowpass filtering

To estimate the intelligibility of the frequency-compressed speech for a deaf person having high-frequency response loss, intelligibility measurements were executed under several lowpass filtering conditions, employing listeners with normal hearing acuity. The cutoff frequencies of the lowpass filters were 1.5, 1.0 and 0.7kHz, with the full frequency bandwidth as reference. As the highest frequency component of the synthesized speech was shifted according to the change of the frequency compression rate, the cutoff frequencies of the full frequency bandwidth were adjusted as 5.0, 4.0, 3.0, 2.5 and 2.0kHz for the frequency compression rates of 1.0, 0.8, 0.6, 0.5 and 0.4, respectively.

3.2.4 Subjects and test materials

Three speakers, one male and two female, read the intelligibility test

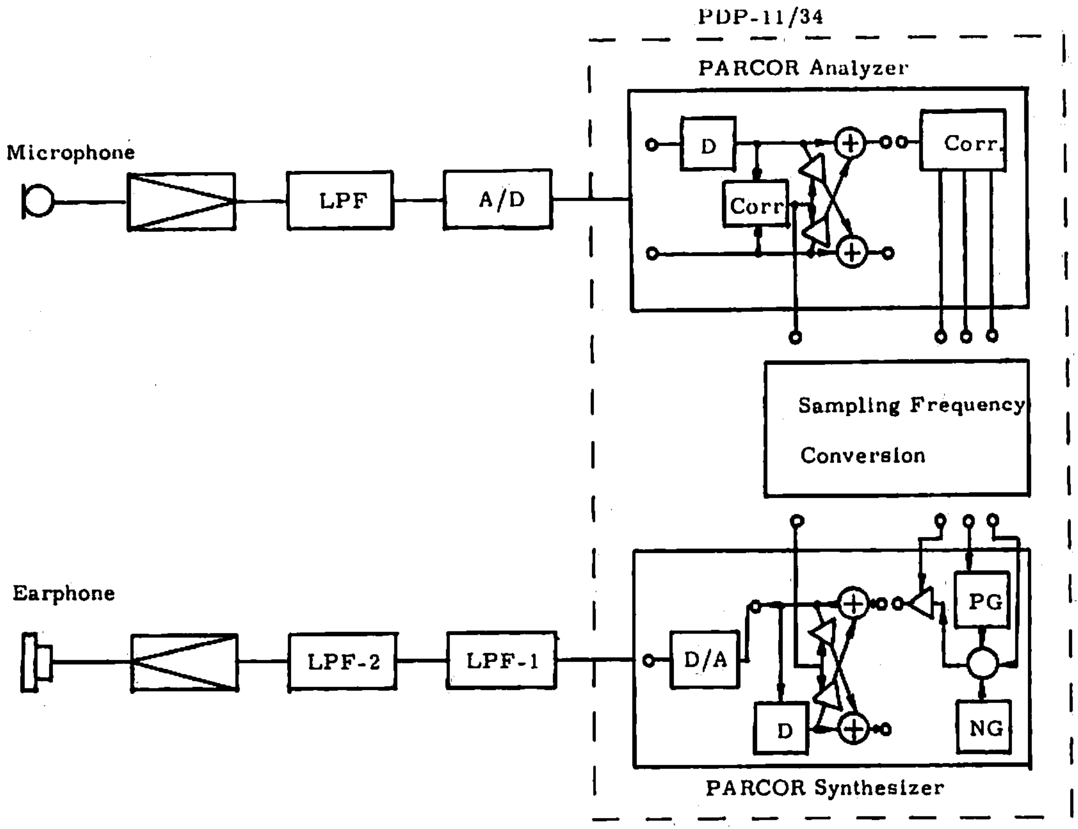

Fig. 2 Block diagram of the experimental equipment

lists containing 100 varieties of Japanese monosyllables. Each test syllable was uttered successively with silent intervals of about 2. 5 seconds. Average speech level at the microphone diaphragm was about 81dB SPL; the microphone was placed at a distance of 15cm from the subject's lips.

Another two subjects, one male and one female, participated in the intelligibility test of the frequency compressed speech as listeners. Listening speech level through the earphone was fixed at 10dB in the orthotelephonic response, which was equivalent to about 73. 5dB SPL in the free sound field. All five subjects have normal ability in speech and hearing. The intelligibility test were held in an anechoic room.

4. Results of Intelligibility Tests

4. 1 Effects of the frequency compression rate

Effects of the frequency compression rate on the articulation score are shown in Figs. 3, 4 and 5 for the three subjects, respectively. These are the results for the synthesized speech with the reduced fundamental frequency of the same rate as the frequency compression. In these figures, the abscissas represent the frequency compression rates and the ordinates represent the decrements of the articulation scores with reference to the non-compressed speech, that is, the compression rate of 100%. The decrements of the articulation scores for vowel and consonant sounds are shown for four kinds of lowpass filtering conditions.

Results show that the consonant articulation scores are decreased monotonously corresponding to the frequency compression rates, while the vowel articulation scores are hardly affected until the compression rates reach 50 – 60% in the case of the full frequency bandwidth, which is denoted as "through". As the frequency bandwidths are reduced to 1. 5, 1. 0 and 0. 7kHz, the decrements of the articulation scores of consonant sounds become small, and in the cases of two female subjects, as is seen in Figs. 3 and 4, increments in the articulation scores are observed. Such increments in the articulation score are prominent in the test condition where the lowpass filtering is 1. 5kHz and the frequency compression rate is around 80 – 60%, although the absolute value of the increment is not significant. The increments of the articulation scores are also observed for female vowel sounds where the test conditions are a lowpass filtering of 0. 7kHz and a frequency compression rate of around 80 – 50%.

In the case of the male subject, on the other hand, the increment in the articulation score is hardly observed for both vowel and consonant sounds, as is seen in Fig. 5, although there are a few increments in the articulation scores only for the frequency compression rate of 80%. It seems that the frequency compression technique is most useful for female speech signals for deaf persons having high-frequency response loss. The preferable compression rates are about 80% for a deaf person with higher cutoff frequency, and about 60% for one with lower cutoff frequency.

4. 2 Effect of the reduction of the fundamental frequency

Similar results for synthesized speech with the original fundamental frequency are obtained and compared with those of the reduced fundamental

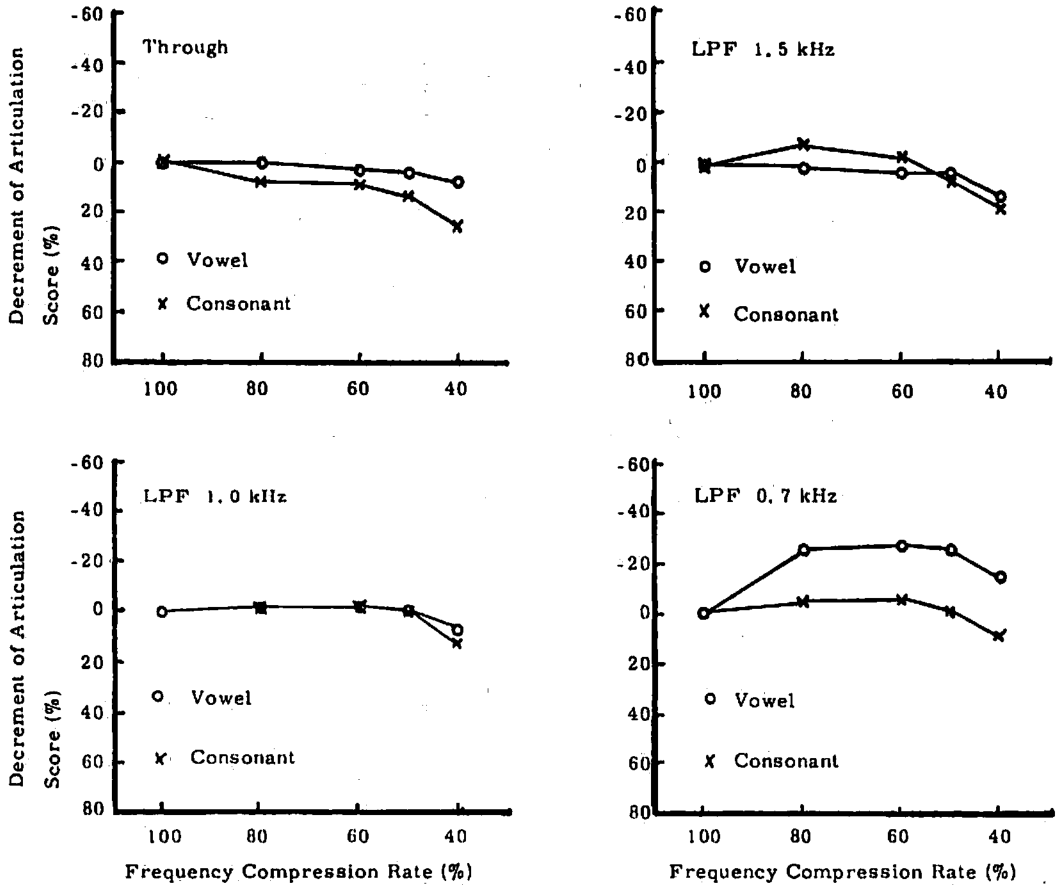

Fig. 3 Effect of the frequency compression rate on the articulation score. Speech sounds are uttered by the female-1 talker and synthesized with the reduced fundamental frequency.

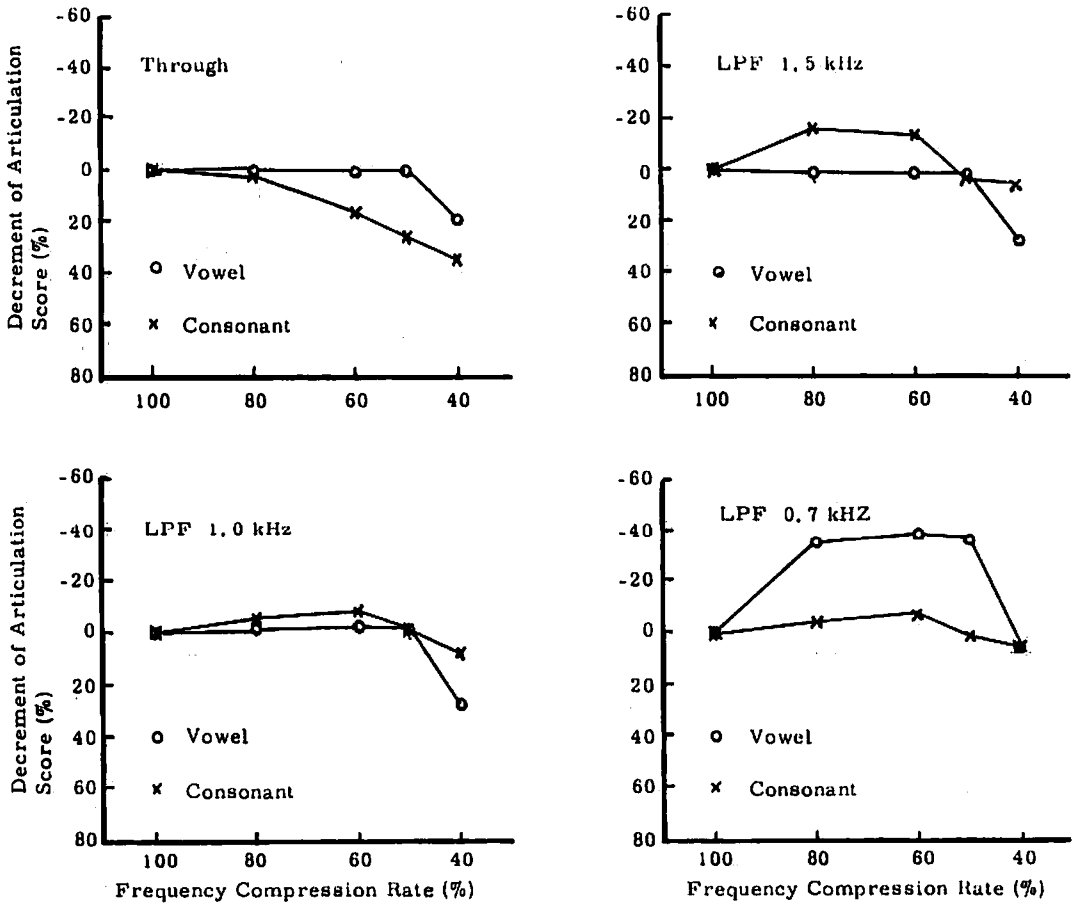

Fig. 4 Effect of the frequency compression rate on the articulation score. Speech sounds are uttered by the female-2 talker and synthesized with the reduced fundamental frequency.

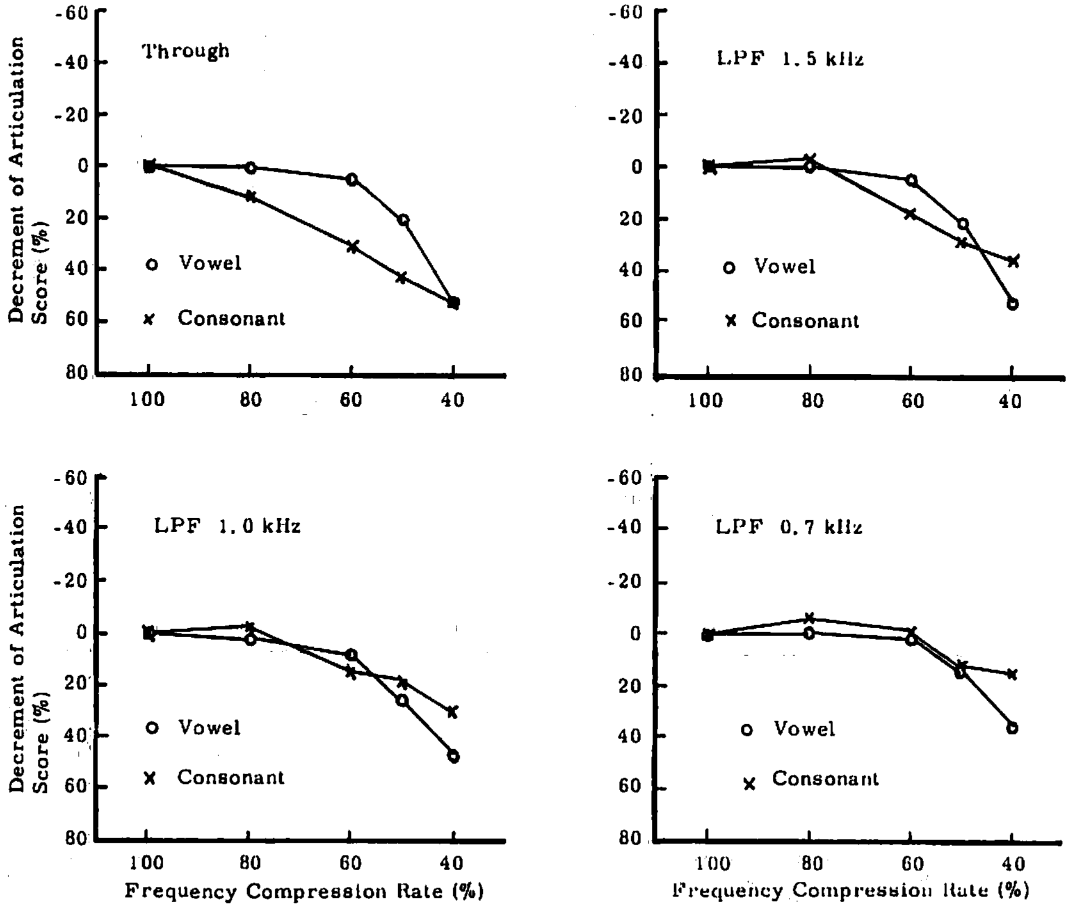

Fig. 5 Effect of the frequency compression rate on the articulation score. Speech sounds are uttered by the male talker and synthesized with the reduced fundamental frequency.

frequency, as shown in Figs. 3, 4 and 5. The differences in the articulation scores for the two kinds of fundamental frequencies are calculated and shown in Fig. 6. In the figure, results for the two female subjects are averaged, because they are very similar, as shown in Figs. 3 and 4. It is seen that the effect of the reduction of the fundamental frequency is significant as the cutoff frequency of the lowpass filtering becomes lower, and is far more significant for the female subjects than the male. It seems that the increment of the vowel articulation score reaches maximum at the frequency compression rate of 60 – 40% for both male and female speech, while that of the consonant articulation score does so at the compression rates of 80% for male and 60% for female.

A few measurements were taken to check the effect of the reduction rate of the fundamental frequency on the articulation score, where the frequency compression rate was fixed at 60%. The reduction rates of the fundamental frequency were varied as 1. 0, 0. 9, 0. 8, 0. 7 and 0. 6. Results are shown in Fig. 7 for the two female subjects.

In lowpass filtering conditions of 1. 5kHz, the maximum increment of the consonant articulation score appears at the reduction rate of about 0. 8. In the narrower lowpass filtering conditions, on the other hand, there is no such convex characteristics, and the maximum increments appear at the same reduction rates as the compression rates for vowel and consonant articulation scores together. In the results for the male subject, no significant increment is found for either vowel or consonant sounds, as is expected from Figs. 5 and 6.

5. Discussion

Summarizing our experiments on the intelligibility of the frequency compressed speech, the vowel articulation scores of the female subjects are much improved by the frequency compression, although the consonant articulation scores of the male and female subjects show a few but not significant improvements and the vowel articulation score is not improved at all. It is noted that the available and effective condition of the frequency compression is not fixed at any single rate, but is varied according to the high-frequency response loss of the deaf person.

For the high-frequency response loss having a cutoff frequency around 1. 5kHz, the compression rate of 80% and the reduced fundamental frequency rate around 0. 9 are the optimum coefficients for female frequency-compressed speech. For the severer high-frequency response loss, such as about 0. 7kHz cutoff frequency, a frequency compression rate of about 60% and a reduced fundamental frequency of the same rate are the optimum ones.

For the male speech, there are a few but insignificant increments for the consonant articulation score and none for the vowels for the variation of the frequency compression rate. The effect of the reduced fundamental frequency is also quite a bit and insignificant either for the vowel or consonant sounds. As a result, it is difficult to choose the appropriate condition of the frequency compression, but it seems that a frequency compression rate of 80% and a reduced fundamental frequency of the same rate are most preferable for the male subject.

As seen in Fig. 5, the effect of frequency compression is zero for the

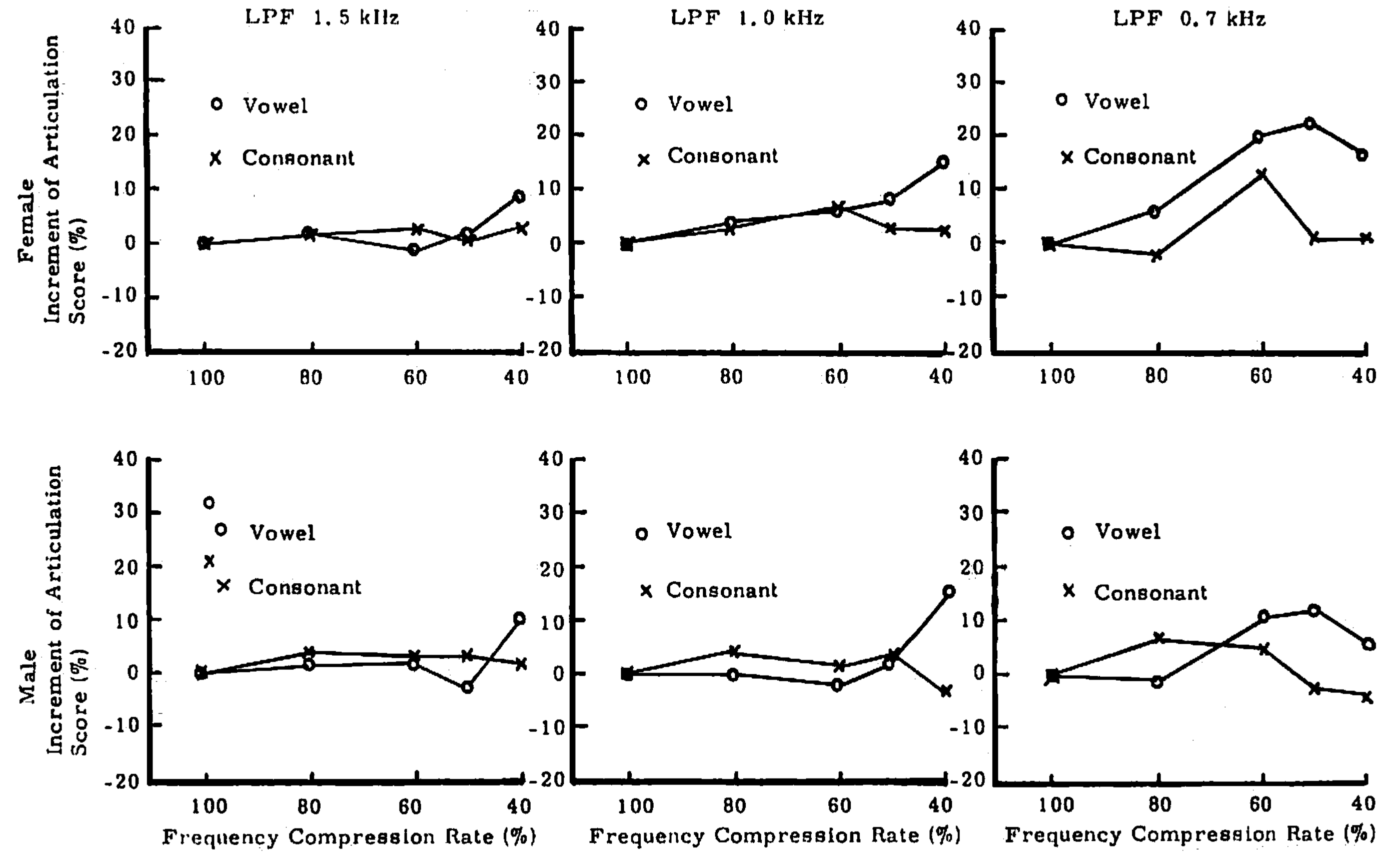

Fig. 6 Effect of the fundamental frequency of the frequency compressed speech on the articulation score. The ordinate represents the difference between the articulation scores for the reduced fundamental frequency and that for the original one.

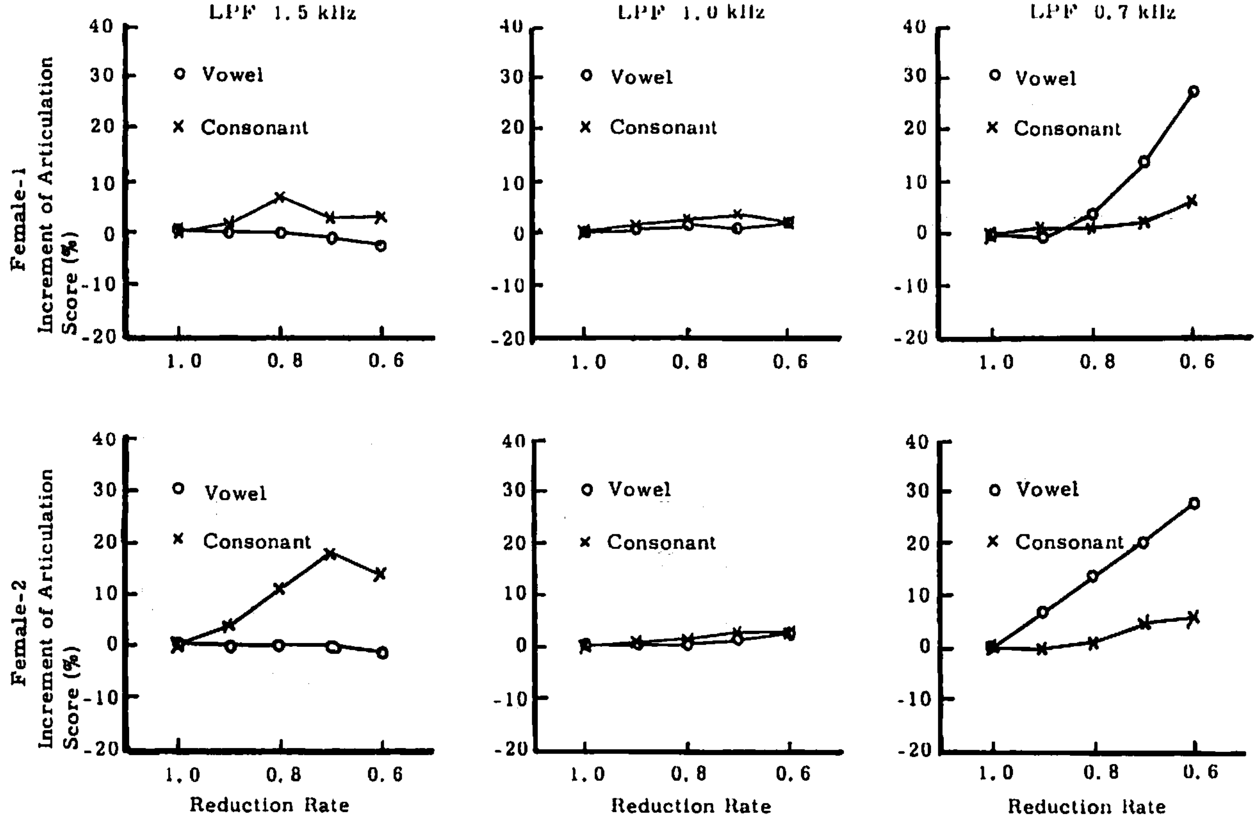

Fig. 7 Effect of the reduction rate of the fundamental frequency of the frequency compressed speech on the articulation score for two female talkers. The ordinate is similar to that of Fig. 6.

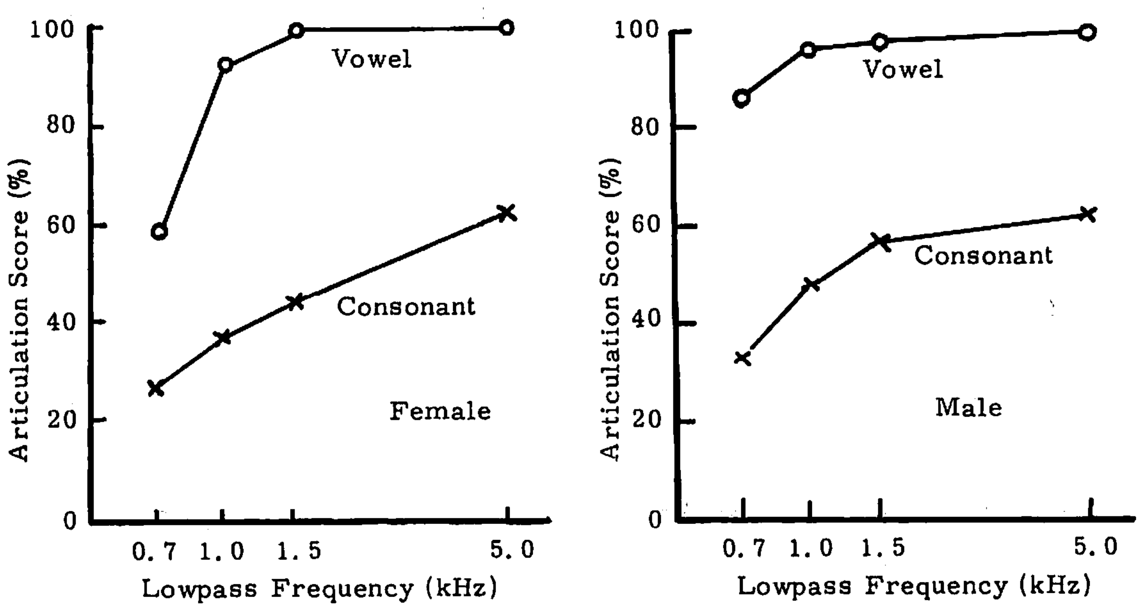

Fig. 8. Articulation score of the synthesized speech with the frequency compression rate of 100 % under the lowpass filtering condition.

male vowels. It seems that such a result is derived from the lower deterioration of the vowel articulation score even in the narrower frequency bandwidth of 0. 7kHz. The articulation scores of the non-compressed speech, that is, the compression rate of 100%, are shown in Fig. 8. Results of the female subjects are averaged in this figure. It is seen that the vowel articulation scores are about 59% for the male and 86% for the female in the case of lowpass filtering of 0. 7kHz. It is supposed that such a difference in the articulation scores may be derived from the differences in vowel formants between male and female. Further study is needed on the lowpass filtering condition narrower than 0. 7kHz for the male speech. It is noted that the consonant articulation scores shown in Fig. 8 are considerably lower than those of the original speech signal. Differences between them range between 5 – 15%. This may result in a few but insignificant improvements in the consonant articulation score. Further technical revisions are needed for the present PARCOR synthesis used in this experiment.

To check the utility of frequency-compressed speech, the intelligibility measurement should be performed by deaf persons. As the hearing loss characteristics of deaf persons vary, however, from person to person, in general it is rather difficult to execute sufficient intelligibility tests for the statistical calculation. We intend to make further measurements of intelligibility using deaf persons, after the intelligibility tests under the simulated response loss of deaf persons have been completed by the subjects with normal hearing.

Acknowledgment

This study was supported in part by a Grant-in-Aid for Scientific Research (Nos. 540003 & 579013) 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) Johansson, B. (1961); A new coding amplifier system for the severely hard of hearing. Proc. 3rd ICA, p. 655.

(2) Pimonov, L. (1962); Coded speech and its application in aids for the deaf. Proc. Speech Comm. Seminar, Stockholm.

(3) Mazor, M. (1977); Moderate frequency compression for moderately hearing impaired, J. Acoust. Soc. Amer., 62, p. 1273.

(4) Drucker, H. (1977); "A review of signal processing techniques for the hearing impaired," in the Application of Signal Processing Concepts to Hearing Aids, P. Yanick, Jr. and S. F. Freifeld (Eds.), Grune and Stratton.

(5) Sekimoto, S., S. Kiritani and S. Saito (1979); Frequency compression techniques of speech using linear prediction analysis-synthesis scheme, Ann. Bull. RILP, 13, p. 133-136.