Ann. Bull. RILP, No. 15, 71 - 90 (1981)

INTELLIGIBILITY OF FREQUENCY COMPRESSED SPEECH AND ITS APPLICATION TO HEARING AIDS

Shuzo Saito, Sotaro Sekimoto and Shigeru Kiritani

1. Introduction

As high frequency sounds carry more intelligibility than low frequency sounds, the sensorineural deaf person with severe hearing loss in the higher frequency region cannot obtain any improvement of perceived speech intelligibility merely by amplification of speech level using conventional hearing aids. Several Techniques of the frequency compression of speech spectrum have been employed to improve the speech intelligibility of sensorineural deaf persons by transposing the higher frequency sounds into the lower frequency region. However, accompanying distortions or noise usually result causing the quality of the frequency compressed speech to deteriorate, thus rendering insignificant any improvement of speech intelligibility.

We applied the PARCOR analysis and synthesis method of speech signal to the frequency compression of speech. Techniques and preliminary results of the intelligibility measurements have been reported previously.[1-4] In this paper, the results on the intelligibility of frequency compressed speech as well as subsequent results on intelligibility measurements are reported. The applicability of frequency compressed speech to hearing aids is also discussed.

2. Test Apparatus

A block diagram of the intelligibility measurement apparatus is shown in Fig. 1. Input speech is uttered into the dynamic microphone located in front of the speaker's lips at a distance of 15 cm. Microphone output was fed to a lowpass filter of 500 Hz cut-off frequency, then sampled at 10,000 Hz and its amplitude digitized into 10 bits. The digital speech signal was then analyzed using the electronic computer PDP-11/34, the PARCOR coefficients of 12 order, excitation source signals of fundamental frequency, source amplitude and the ratio of voiced and unvoiced source amplitudes.

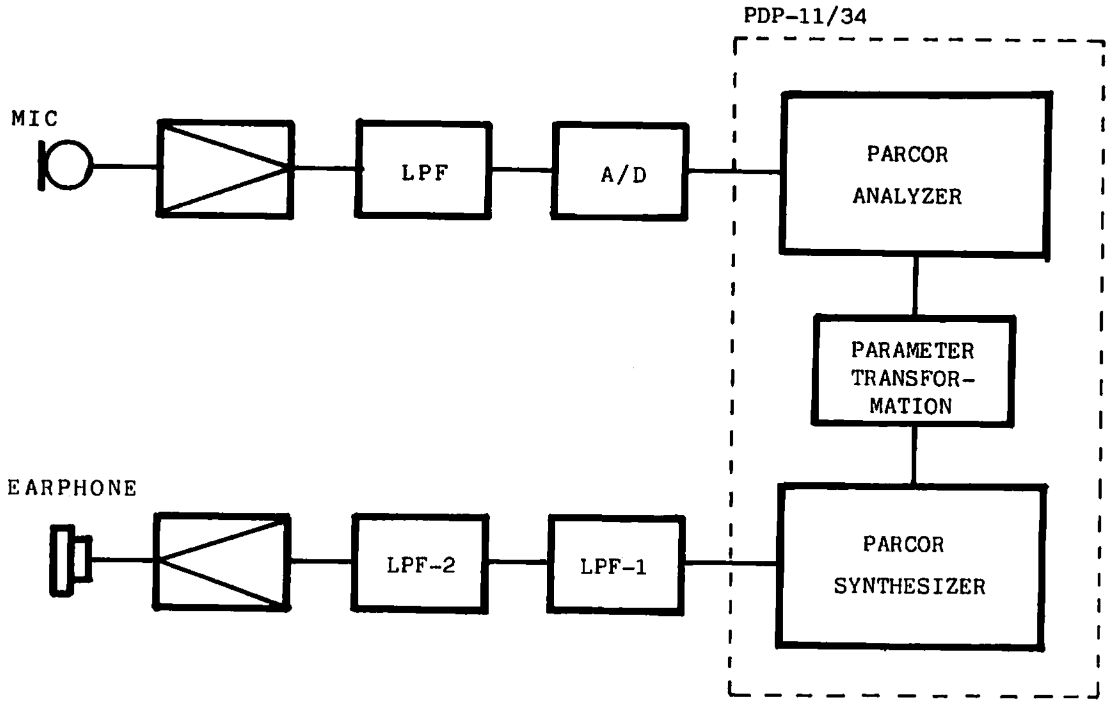

Fig. 1 Block diagram of the intelligibility measurement apparatus.

These analyzed feature parameters were used to synthesize frequency compressed speech in which the frequency spectrum envelope of speech was shifted to a frequency region lower than that of the input speech. Such frequency compression processing was performed by the reduction of the sampling frequency of the PARCOR synthesizer from that of the analyzer. As the frame interval of the PARCOR synthesizer was set at 5 milli-seconds, the same as in the PARCOR analyzer, the utterance speed of the synthesized speech coincided with the input speech, although the numbers of speech samples in each frame interval differed between the input and the synthesized speeches. As the fundamental frequency of the PARCOR synthesizer is able to adjust freely, in addition to the frequency compressed speech having the original fundamental frequency, the frequency compressed speech having reduced fundamental frequency in proportion to the frequency compression rate was synthesized and used in the intelligibility measurement. Modifications of the feature parameters and the PARCOR synthesis were made with the aid of the computer PDP-11/34.

The PARCOR synthesizer output was D/A converted and passed through two lowpass filters. The cut-off frequency of the LPF-1 was adjusted in accordance with the sampling frequency of the PARCOR synthesizer to eliminate the decoding noise.

The LPF-2 was used to simulate the high frequency response loss characteristics found in sensorineural deaf persons. To assess the effect of frequency compressed speech on sensorineural deaf persons, intelligibility measurement should be made by listeners who are deaf. However, actual high frequency response loss characteristics of deaf persons vary among individuals so the effect of the frequency compressed speech on speech intellibility will also vary, making it rather difficult to estimate statistically the availability of frequency compressed speech from the intelligibility measurement made by deaf persons. In this study, intelligibility measurement was performed by listeners with normal hearing, and the high frequency response loss characteristics were simulated by several characteristics of the lowpass filter-2. The cut-off frequencies of the lowpass filter ranged between 500 - 1,500 Hz and two kinds of slopes, that is, steep and gentle, were used for each cut-off frequency. It is noted that these cut-off characteristics of the lowpass filter-2 were selected as first order approximations, but they are not the typical or average characteristics of deaf persons.

The output of LPF-2 was fed to the dynamic receiver DR-305 and the real ear response of the overall receiving system was within 3 dB in the 5,000 Hz frequency band.

3. Measuring Method

Intelligibility of the frequency compressed speech was measured under the various test conditions of frequency compression rate, fundamental frequency, cut-off frequency, slope of the low-pass filter-2 among others.

The frequency compression rates of the synthesized speech were 100, 80, 60, 50 and 40%. In the case of 100% compression rate, the frequency spectrum was not compressed, but held the same as the speech input. In 40% compression rate the upper limit frequency of the frequency spectrum reduced to 2,000 Hz (=5,000 Hz x 0.4).

As for the fundamental frequency of the frequency compressed speech, two test conditions were used as described in the previous section, i.e., one is the same as the original and the other is reduced at the same rate as the frequency compression rate.

The cut-off frequencies of the lowpass filter-2 used for the intelligibility measurement were 1,500, 1,000, 700 and 500 Hz, and the slopes of the filter were both more than 120 dB per octave and 24 dB per octave. Intelligibility measurements were performed mainly by lowpass filter with steep slope, and the filter with gentle slope was used in a supplementary capacity.

In the intelligibility measurement, three subjects, one male and two female, were engaged as speakers and two subjects, one male and one female, as listeners. All had normal speech utterance and hearing acuity, and were trained in the intelligibility measurement for about one month.

Speech materials were 100 Japanese monosyllables uttered separately in an anechoic room. The speech level of the frequency compressed speech presented to the listeners through the dynamic earphones was about 73.5 dB in the reference condition, that is, the system gain of the intelligibility measurement circuit was adjusted to 10 dB in orthotelephonic reference gain.

4. Intelligibility of the Frequency Compressed Speech

4.1 Effect of the frequency compression rate

Effects of the frequency compression rate on intelligibility are shown in Figs. 2 and 3 as the results for the two female speakers were very similar, they were averaged and shown in Fig. 2. Figure 3 shows the results for the male speaker. In these figures, the abscissae represent the frequency compression rate and the ordinates represent the increment of the articulation score of the test monosyllables with reference to noncompressed speech. The increments of the vowel and consonant articulation scores are shown separately.

As seen in Fig. 2, the consonant articulation scores improved significantly at frequency compression rates of 80 - 60%, lowpass filtering of 1,500 Hz, while the vowel articulation scores improved very significantly at frequency compression rates of 80 - 50%, lowpass filtering of 700 Hz.

Figure 3 shows that the consonant articulation scores were lowered by frequency compression, except at 80% compression of all LPF conditions, although improvement of articulation scores of the 80% compressed speech is insignificant. The vowel articulation scores improved at 80 - 60% frequency compression of lowpass filtering of 500 Hz. The improvement is significant, but not as much as in female speech.

From these results, it can be concluded that the effect of the frequency compression rate on the articulation score is affected by the sex difference of speech input as well as by the cut-off frequency of the lowpass filtering. In female speech, improvement of articulation score is very significant, but very little improvement is observed in male speech. Improvement of the articulation score for vowel sounds in female speech is prominent in lowpass filtering of 700 Hz, and for consonant sounds is superior in lowpass filtering of 1,500 Hz. It is noted that the articulation scores described here are those of speech input with reduced fundamental frequency. Effect of the fundamental frequency of frequency compressed speech will be described in the following section.

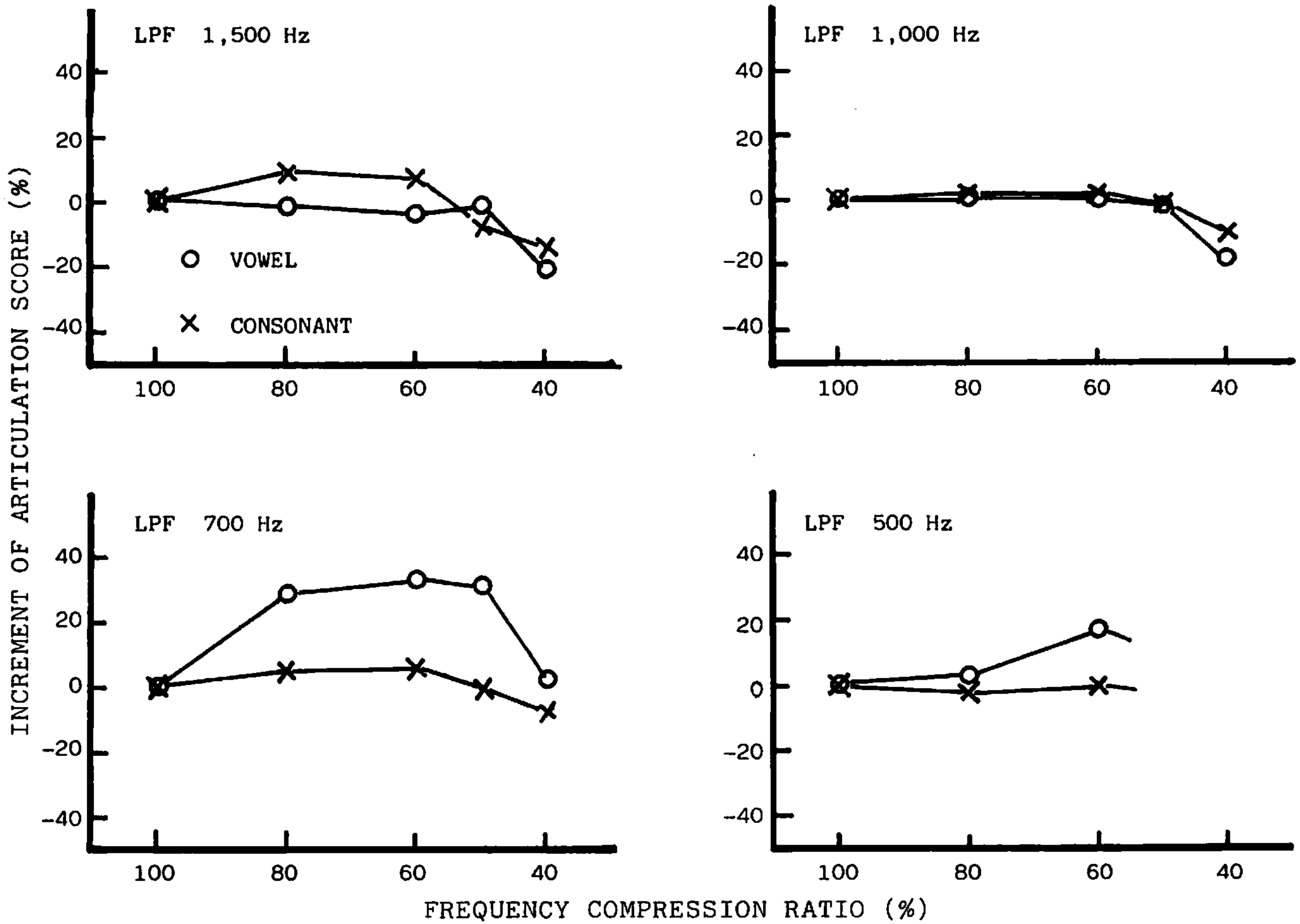

Fig. 2 Effect of frequency compression on the articulation score of female speech.

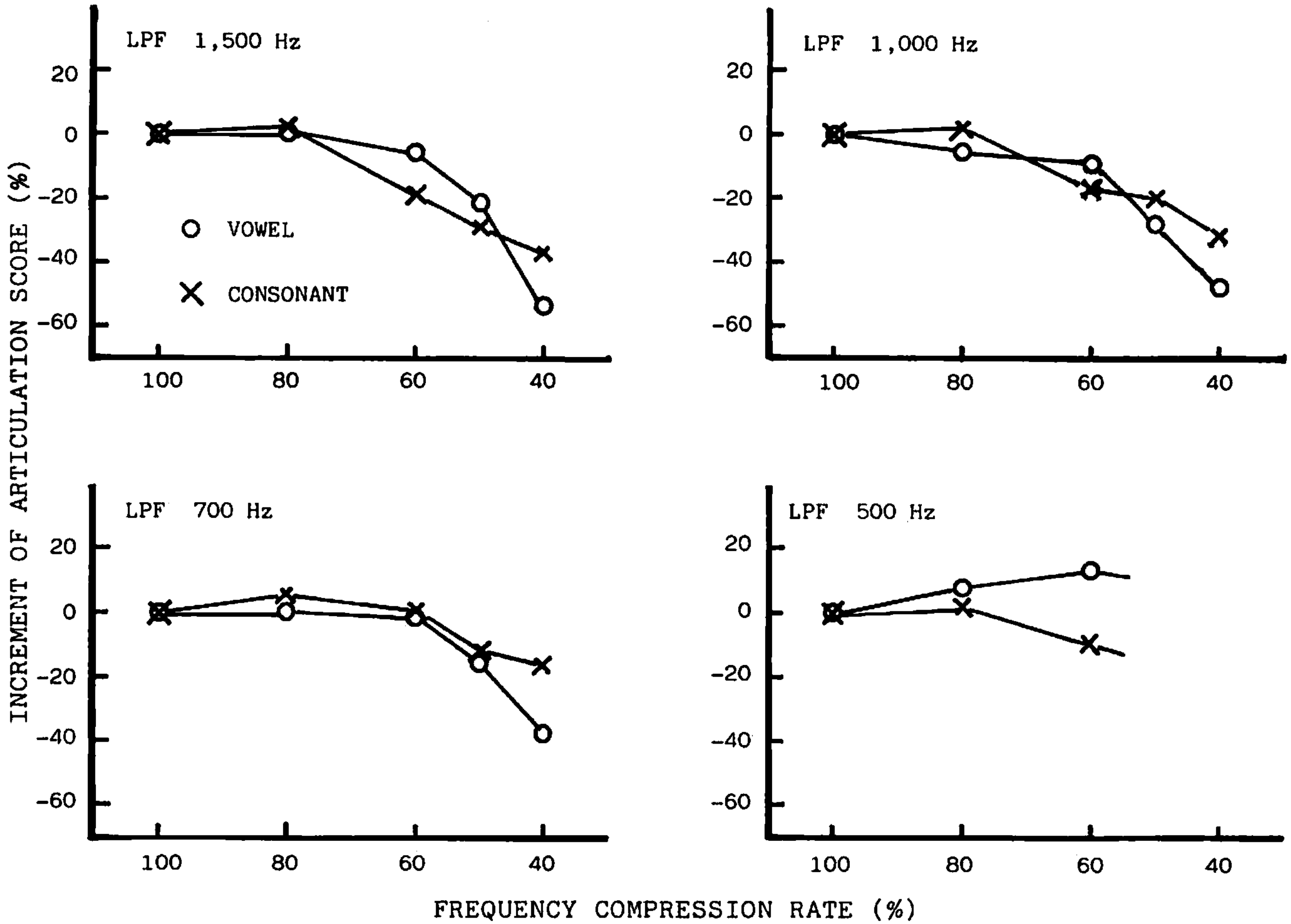

Fig. 3 Effect of frequency compression on the articulation score of male speech.

4.2 Effect of the fundamental frequency

Results of intelligibility measurement conducted for frequency compressed speech with the original fundamental frequency are shown in Fig. 4. It is seen that the increment of the articulation score becomes maximum at the compression rate of 80% for frequency compressed speech with the original fundamental frequency. Comparing Fig. 4 with Fig. 2, it appears that the effect of the reduction of the fundamental frequency of frequency compressed speech on the vowel articulation score is prominent when the compression rates are in the range of 60 - 40%. The increment of the articulation score reaches maximum at 80 - 50% compression rate. The increment of the consonant articulation score is similar though smaller, but the decrement by the frequency compression is larger, so that the resulting increment of the consonant articulation score becomes insignificant.

4.3 Effect of the slope of the lowpass filter

In the intelligibility measurements of frequency compressed speech shown in Figs. 2 - 4, the cut-off slope of the lowpass filter to simulate the high frequency response loss of the deaf person was steeper than 110 dB per octave. Results using the gentle slope lowpass filter are shown in Fig. 5. It is seen that the improvement of intelligibility by frequency compression is very similar to that shown in Figs. 2 and 3, except that the effect of the cut-off frequency of the lowpass filter is transformed and the cut-off frequency of the gentle slope filter corresponds to the higher cut-off frequency of the steep slope filter. The articulation score measured by the gentle slope lowpass filter was compared with that measured by the steep slope filter and the articulation equivalent cut-off frequency of the steep slope filter estimated. Figure 6 shows the results for male and female speech, vowel and consonant articulation scores. It can be concluded that the cut-off frequencies of 1,000 and 500 Hz of the gentle slope filters are equivalent to the cut-off frequencies of 1,180 and 760 Hz of the steep slope filters, respectively, although the correspondences of the cut-off frequencies as shown in Fig. 6 diverged somewhat in the lower cut-off frequency.

The cut-off frequencies shown in Fig. 5 were corrected, then

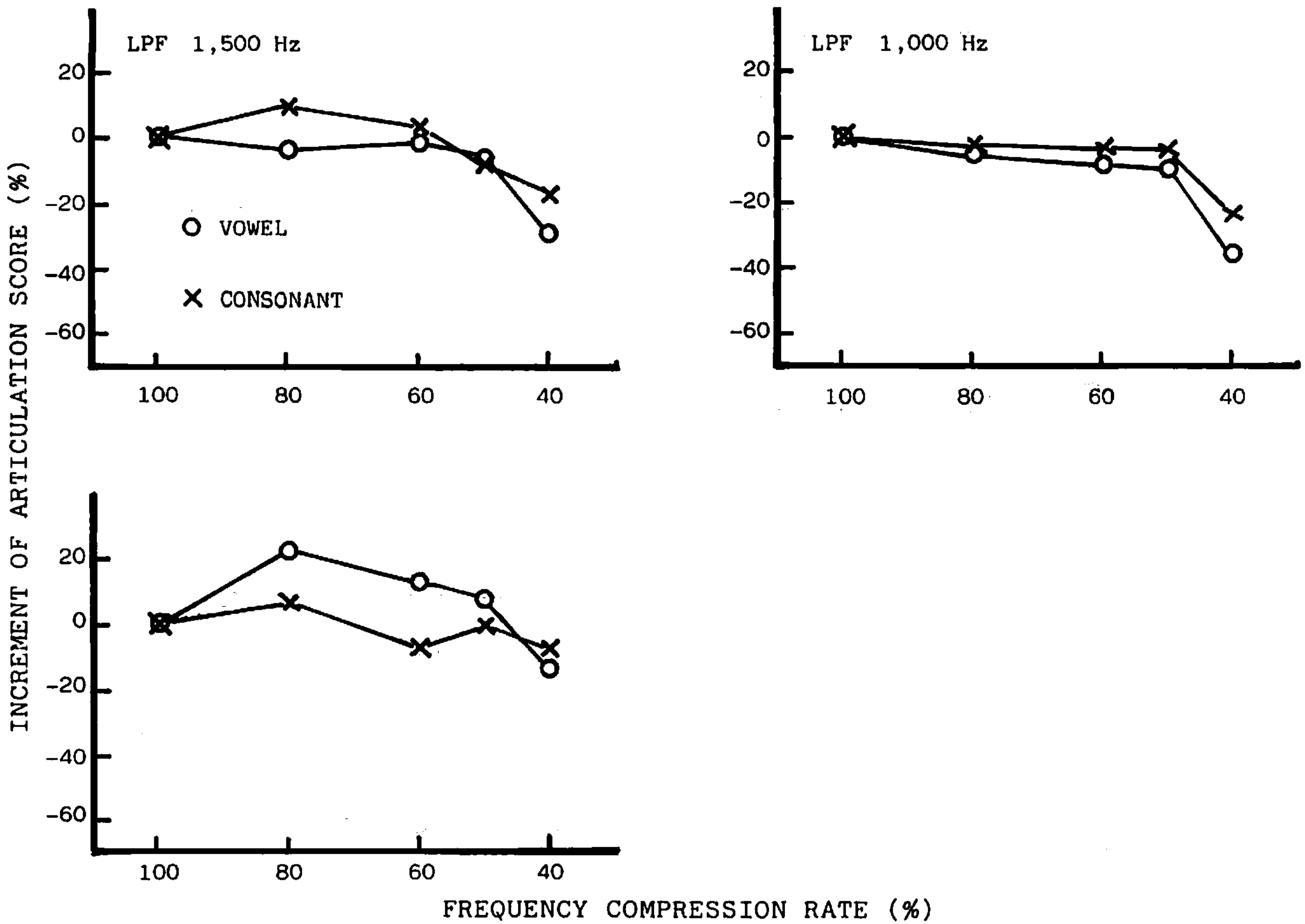

Fig. 4 Effect of frequency compression on the articulation score of female speech having the original fundamental frequency.

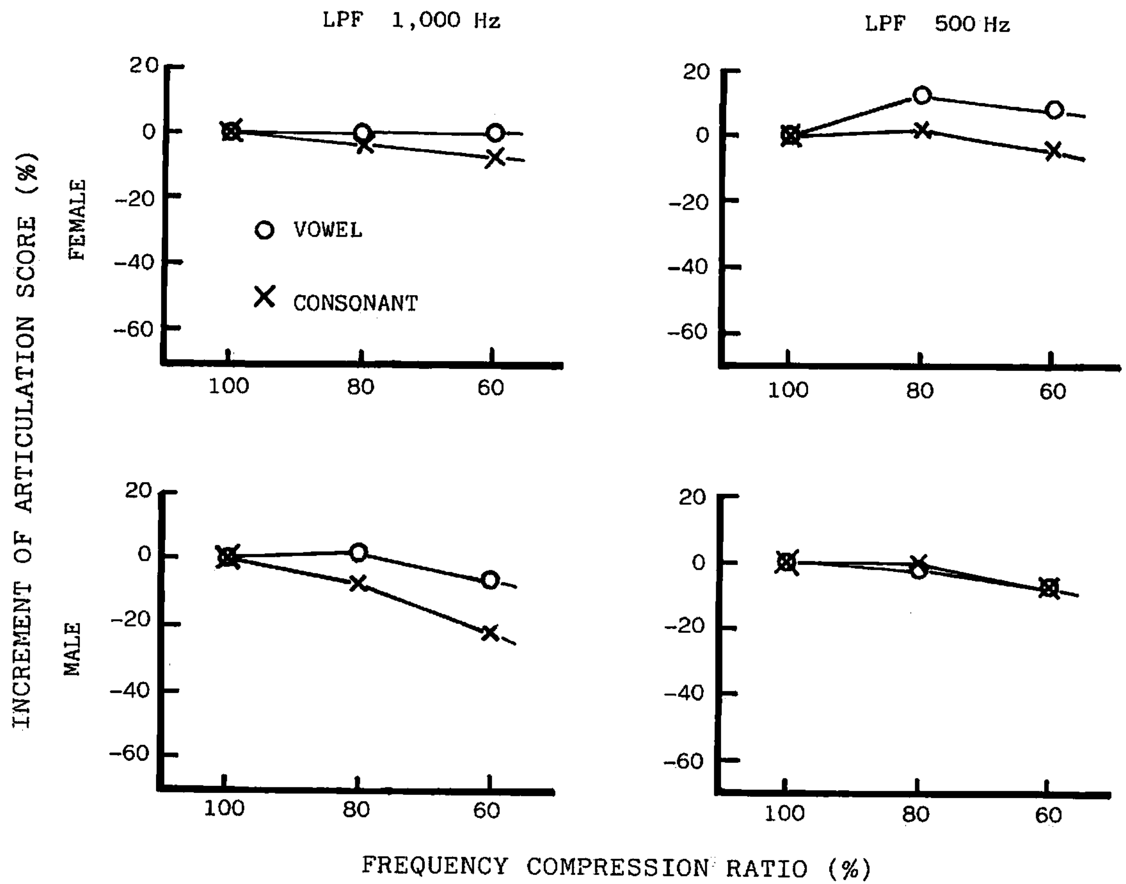

Fig. 5 Increment of the articulation score of frequency compressed speech passed through the gentle slope lowpass filter.

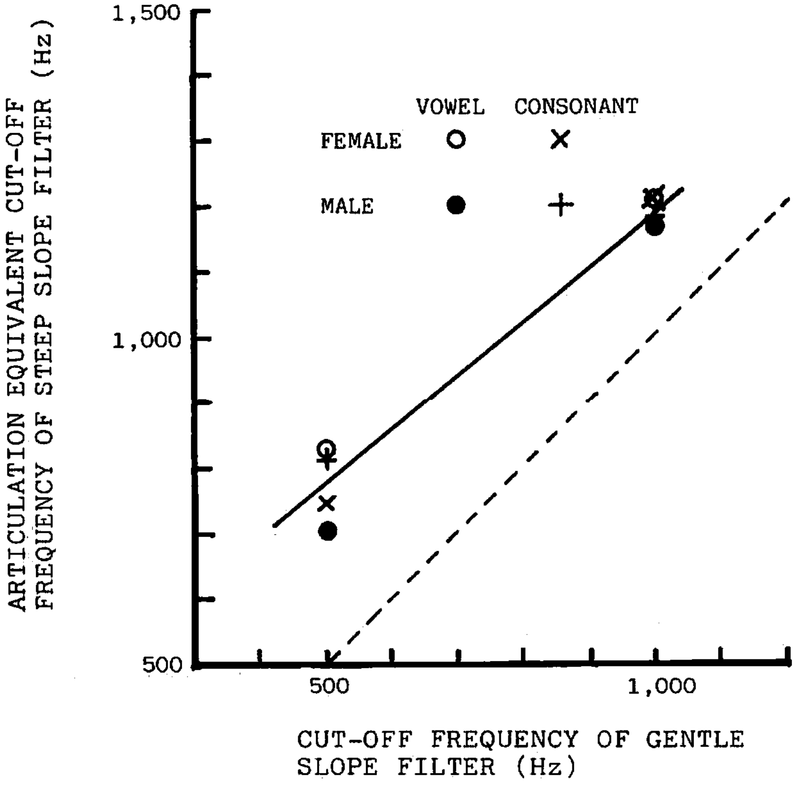

Fig. 6 Estimated cut-off frequency of the gentle slope lowpass filter.

compared with the results of Figs. 2 and 3. It can be concluded that the effect of relaxation of the cut-off slope on the articulation score of frequency compressed speech is equivalent to the upwards shift of the cut-off frequency of the steep slope lowpass filter.

5. Perceptual Confusion of Frequency Compressed Speech

To analyze the articulation scores of frequency compressed speech in detail, perceptual phoneme confusion matrices were derived using the intelligibility measurement data described above. The 100 Japanese monosyllables used are composed in the form of consonant + vowel (CV). The phoneme confusion matrices were calculated for the five Japanese vowels /i/, /e/, /a/, /o/, /u/ and the 26 kinds of Japanese consonants as shown in Table 1.

(1) Confusion of vowel sounds

To show the results of the confusion analysis concisely, the diagonal elements of the confusion matrices, that is, the articulation scores of the five vowels, are illustrated in Fig. 7. These are the results for frequency compressed speech in the case of the lowpass filtering of 700 Hz.

It is seen that in female speech the articulation scores of vowels /a/ and /o/ are strikingly low and confused mainly as /e/ at the frequency compression rate of 100%, but they are greatly improved by frequency compression of 80 - 50%. The articulation score of the vowel /i/, on the contrary, is decreased gradually by frequency compression, and those of vowels /u/ and /e/ are less affected by frequency compression. On the other hand, in male speech improvement of the articulation score by frequency compression is very slight and vowel /e/ is confused mainly with /o/ and vowel /u/ is confused mainly with /i/.

(2) Confusion of consonant sounds

The Japanese consonants are classified by the manner and place of articulation as shown in Table 1. Confusions of the consonant sounds are calculated between such groups as the manner and place of articulation.

The articulation scores of the voiced and unvoiced consonant groups are shown in Fig. 8. It appears that the frequency compression hardly affects the voiced-unvoiced discrimination in lowpass filtering of 1,500 Hz. Similar articulation scores of the conso-

MANNER / PLACE	VOICED	UNVOICED
LABIAL	b, bj, m, mj, w	p, pj
DENTAL	d, dz, n, nj	t, ts, tʃ, s
FLAPPED	r, rj	
PALATAL	dʒ, j	ʃ, hj
VELAR	g, gj	k, kj
GLOTTAL		h

Table 1 Classification of Japanese consonants

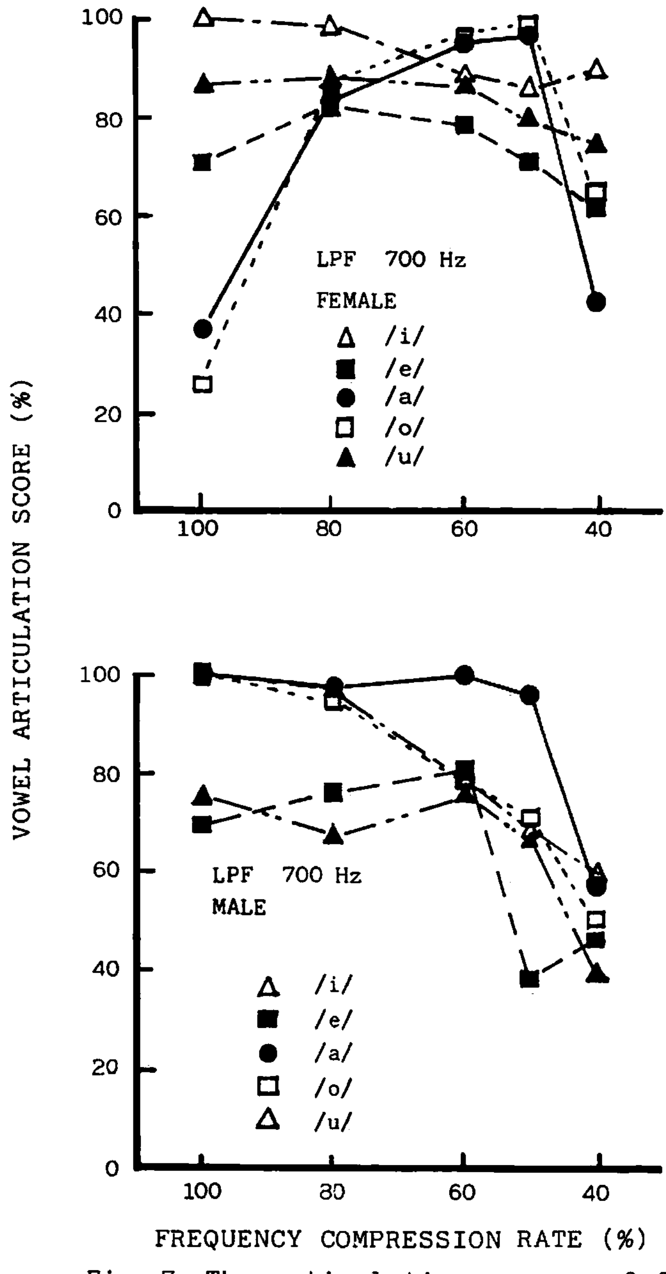

Fig. 7 The articulation scores of five Japanese frequency compressed vowels.

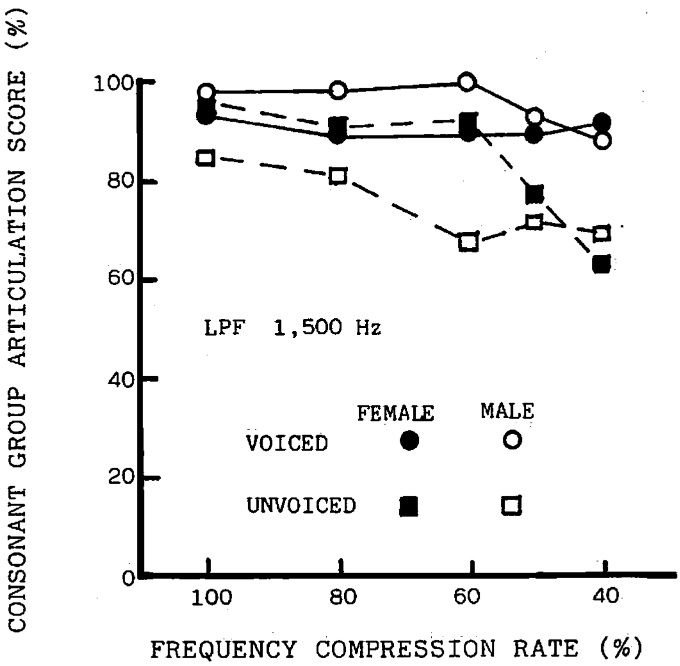

Fig. 8 The articulation scores of the voiced and unvoiced consonant groups, expressed in relation to the frequency compression rate.

nant groups classified by place of articulation are shown in Fig. 9. It is seen that the articulation score of the labial consonant group is improved significantly by frequency compression of 80 - 60% and those of the dental and velar consonant groups are also improved by frequency compression of 80% in the case of female speech. Similar improvement by frequency compression is observed at the 80% compression rate for the labial consonant group in male speech.

6. Application of Frequency Compressed Speech to Hearing Aids

The results of the intelligibility measurement on the frequency compressed speech described in the preceding sections can be summarized as follows.

(1) The optimum frequency compression rate for improving speech intelligibility for a deaf person depends on his hearing loss characteristics and the fundamental frequency of the input speech signal. For example, if the cut-off frequency of the high frequency loss characteristics is higher than 1,000 Hz, the frequency compression rate of 80% is preferable, and when the cut-off frequency is lower than 1,000 Hz, the frequency compression rate of 60 - 70% is better. In addition, it is preferable to further intensify the frequency compression rate by 10% for female speech input.

(2) It is necessary to reduce the fundamental frequency of frequency compressed speech in proportion to the frequency compression rate for female speech input. Such a reduction of the fundamental frequency is not necessary for male speech input.

(3) The gentle cut-off slope of the high frequency loss characteristics is able to compensate by removing the cut-off frequency of the steep cut-off characteristics.

Hearing loss characteristics of the sensorineural deaf vary among individuals and, as a matter of course, the fundamental frequencies of the input speech signals uttered to a deaf person vary widely. Thus when constructing hearing aids employing frequency compressed speech the following must be taken into consideration.

(a) The frequency compression rate can be adjusted in advance corresponding to the high frequency loss characteristics of the individual deaf person.

(b) The adjusted frequency compression rate can modify to

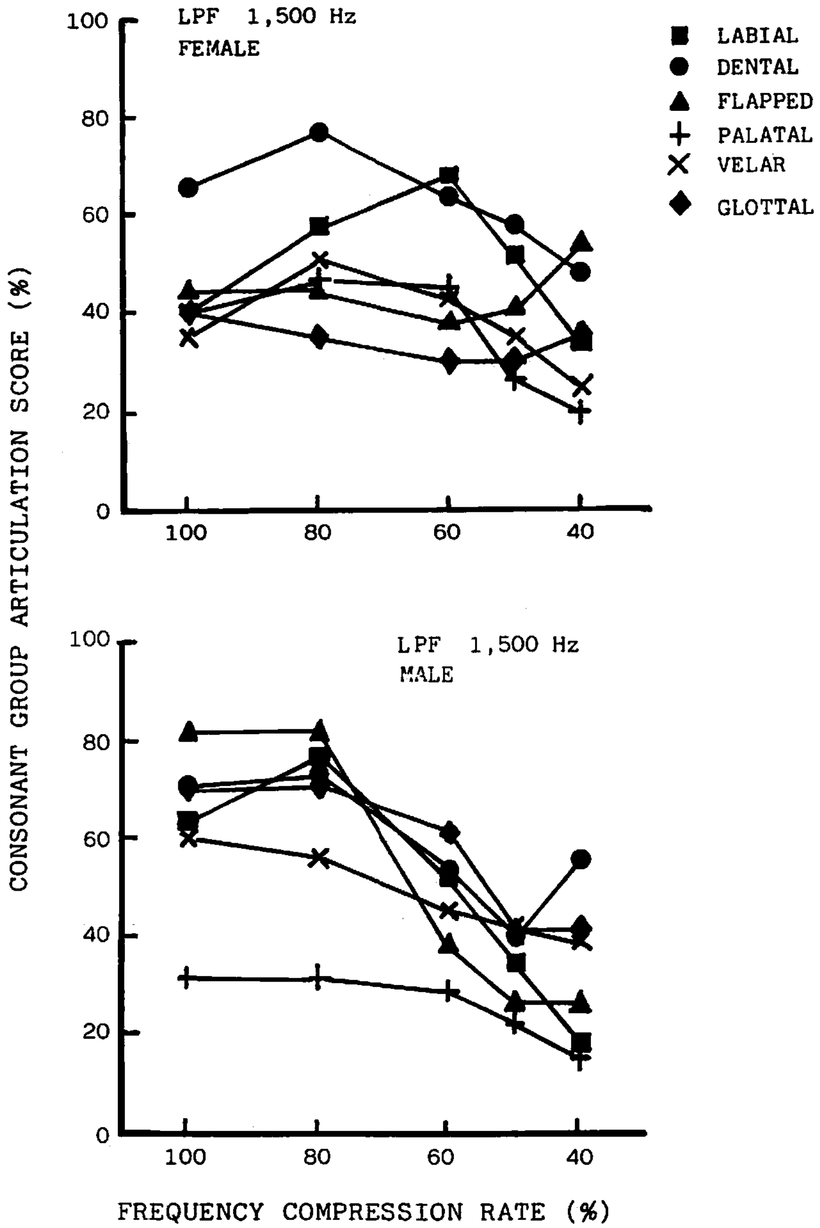

Fig. 9 The articulation scores of the consonant groups classified by place of articulation, expressed in relation to the frequency compression rate.

correspond to the average value of the fundamental frequency of speech input to hearing aids. The average fundamental frequency is analyzed automatically in the hearing aids. To put it concretely and simply, the frequency compression rate is intensified up to 10% when the average fundamental frequency exceeds the threshold value of 180 Hz.

(c) The fundamental frequency of frequency compressed speech is able to reduce in proportion to the frequency compression rate when the average fundamental frequency exceeds the threshold value of 180 Hz.

These functions are implemented by electric circuit, as shown in Fig. 10. This circuit is inserted between the PARCOR analyzer and the synthesizer, as shown in Fig. 1. The analyzed PARCOR coefficients and the excitation source signals are set at the registers in every frame period. The frequency compression rate is set in the FCR controller in advance through the FCR adjust terminal and is modified by the average value of the fundamental frequency of the speech input. The fundamental frequency of the synthesized speech is also modified by the averaged fundamental frequency.

7. Conclusion

The intelligibility of frequency compressed speech was measured employing listeners with normal hearing acuity and lowpass filters. The construction of hearing aids of frequency compressed speech is discussed taking the results into consideration.

As the hearing loss characteristics of the sensorineural deaf differ from person to person, the results described in this paper are applicable only to a limited number of deaf persons. Further investigation is needed employing deaf persons to measure the intelligibility of frequency compressed speech. In a preliminary study of the intelligibility of frequency compressed words, it has been reported that the intelligibilities of one-fourth of the deaf persons who participated improved significantly with the use of frequency compressed speech as described in this paper.[5]

Further studies are necessary to collect more data on the intelligibility of frequency compressed speech measured by deaf persons and to elucidate the applicability of frequency compressed speech to hearing aids.

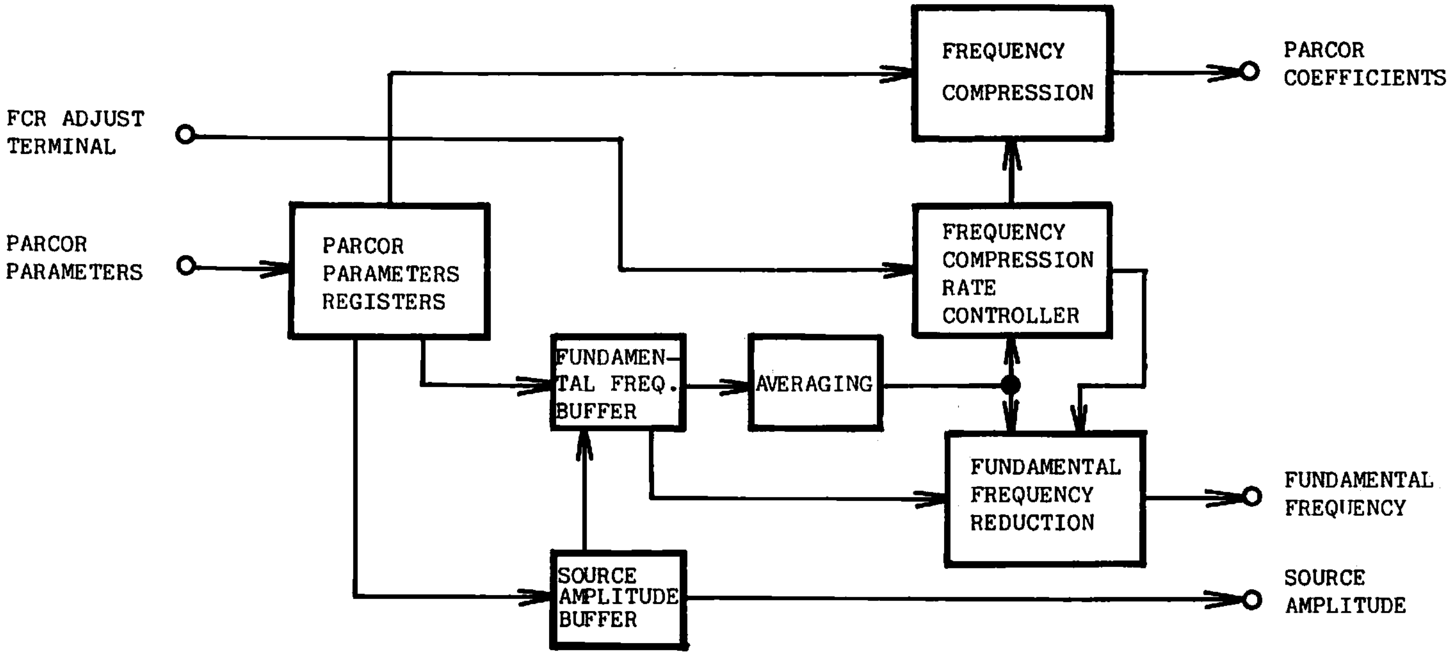

Fig. 10 Block diagram of the parameter transformation circuit used in hearing aids for sensorineural deaf persons.

Acknowledgment

This study was supported in part by Grants-in-Aid for Scientific Research (Nos. 0054003 & 56790019) of the Japanese Ministry of Education, Science and Culture and by a Grant-in-Aid for Research on Computed Ear of the Japanese Ministry of Science and Technology.

References

1) 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.
2) Sekimoto, S. and S. Saito (1980); Nonlinear frequency compression speech processing based on the PARCOR analysis-synthesis technique, Ann. Bull. RILP, 14, 65-72.
3) Sekimoto, S., S. Kiritani and S. Saito (1980); Intelligibility of frequency compressed speech in lowpass filtered condition, Ann. Bull. RILP, 14, 181-193.
4) Sekimoto, S., S. Kiritani and S. Saito (1980); Perceptual phoneme confusion of frequency compressed speech, Ann. Bull. RILP, 14, 195-200.
5) Manabe, T., K. Kiba, T. Tateishi and K. Yonemoto (1980); Intelligibility of the frequency compressed speech processed by PARCOR synthesis in sensorineural deaf persons, Audiology, Japan, 23, 543-544.