Ann. Bull. RILP (1982) No. 16, 89-94

EFFECTS OF FORMANT PEAK EMPHASIS ON VOWEL INTELLIGIBILITY IN FREQUENCY-COMPRESSED SPEECH

Sotaro Sekimoto

1. Introduction

In a previous paper we proposed a frequency lowering technique for hearing aids[1]. It was verified that the technique did improve vowel intelligibility effectively in case of simulated hearing loss[2]. Since it was, however, not apparent why the technique improved the intelligibility, the purpose of the present study was to clarify the mechanism of the perception of frequency-compressed speech.

In the present study, we examine the identification cues for frequency-compressed vowels. It is well known that formants play an important role in identifying vowels. Therefore, the effects of formant peak emphasis on the intelligibility of frequency-lowered vowels were investigated. We also conducted another experiment in order to explore the normalization process of the frequency scale, and the results are presented in other part of this issue[3].

2. Method

2.1 Measurement

Five steady state Japanese vowels uttered by three speakers, a male and two female adults, were recorded in a soundproof room. The speech materials were processed by the method to be described below, and then subjected to the intelligibility test, which was carried out under the following conditions:

(i) Bandwidth ratio

The bandwidths of each formant were 30%, 50%, 70%, and 100% (not reduced) of the original.

(ii) Frequency compression rate

The spectrum envelope was linearly compressed to 60% of the original frequency scale.

(iii) Cutoff frequencies of the low-pass filter

A low-pass filter was used in order to simulate the frequency response of the hearing impaired. The cut-off frequencies were 0.5 kHz and 0.7 kHz.

The processed speech stimuli were presented in random order to a male subject with normal hearing. The subject was requested to identify each stimulus as one of the five Japanese vowels. The tests were carried out in a soundproof room using a headphone for one ear. The presentation sound level was fixed at 10 dB in the orthotelephonic response.

2.2 Processing of the Speech Wave

Fig. 1 shows the flow of the speech processing system. The formant peak emphasis was executed in the following way: the utterance was low-pass filtered

at 5 kHz and sampled at 10 kHz through a A/D converter; then analyzed by a PARCOR analyzer in order to extract the LPC coefficients (a_0-a_{12}) and the fundamental frequency (F_0) through a Hamming window 30 msec in length. The analysis was repeated at 5 msec intervals. Formant frequencies and bandwidths were calculated from the polynomial of the LPC coefficients. Bandwidths were reduced at the specified formant bandwidth reduction rate, whereas formant frequencies were kept unchanged. The LPC coefficients (a'_0-a'_{12}) were reconstructed from the modified formant frequencies and the bandwidths. These LPC coefficients were converted into PARCOR coefficients by the step-down recursion process. The PARCOR coefficients were transferred to the PARCOR synthesizer to produce the frequency-lowered speech material. Frequency lowering was achieved by lowering the sampling frequency of the synthesizer.

3. Results

Fig. 2 shows the relation between vowel intelligibility and the bandwidth ratio, where the cutoff frequency of the low-pass filter was set at 0.7 kHz. In the figure, a heavy line shows the vowel articulation score, and light lines show the identification rate for the respective vowels.

For the female-1 voice, shown in Fig. 2 (a), the vowel articulation score increased when the formant bandwidth was narrowed. This increase in the score seems to have been caused by a notable improvement in the identification rate for /e/.

From the results of the female-2 voice and the male voice, shown in Fig. 2 (b) and 2 (c), respectively, we can also see an increase in vowel intelligibility scores. However, the vowel identification score does not increase for /e/, while it does for /a/, /i/ and /o/.

Fig. 3 shows the intelligibility scores when the cut-off frequency of the low-pass filter was set at 0.5 kHz. The frequency compression rate was 60%. The figure shows that the vowel intelligibility score increased for all speakers except female-2. However, the vowel for which the identification score increased varied with the speakers. With the female-2 voice, the vowel identification score increased for some vowels, although vowel intelligibility did not improve.

4. Remarks

In the present study, the effect of formant peak emphasis on vowel intelligibility for frequency-compressed speech was explored under the low-pass condition which simulates the characteristics of the hearing impaired. It was revealed that formant peak emphasis is effective in increasing vowel intelligibility scores for frequency-compressed speech. The range of the bandwidth ratio with which the vowel intelligibility scores increased was found to be from 30% to 70% for the speech materials used. As the identification score varied depending upon the vowels, the vowels were divided into two groups; for one group the identification score increased with formant peak emphasis. Moreover, the results were influenced by the low-pass condition and speaker differences. It was felt that these variations were caused by

the difference in the number of formants within the pass-band which was determined by the identity of the vowel, the low-pass condition and the speakers. These results imply that formants serve as a cue to the identification of vowels even in formant-compressed speech. However, only one formant sometimes falls into the pass-band for /i/. In such a case, it seems that the function of the formant as a perceptual cue for identifying frequency-compressed vowels is somewhat different from that for normal speech. Thus, further study is indicated.

Amplitude processing is left as one of the pending problems regarding hearing aids of frequency-lowering type. In the ordinary hearing aid, such as the compression amplifier type, the use of the identification cue is not intended. From the results of the present study, it can be suggested that speech perceptual cues in amplitude processing would be useful for the hearing impaired. Hearing aids based on the PARCOR speech analysis-synthesis technique proposed by the authors is advantageous, since the method can make use of such cues more easily than conventional ones.

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., S. Kiritani and S. Saito (1980); Intelligibility of frequency compressed speech in low-pass filtered condition, Ann. Bull. RILP, 14, 181-193.
2. Sekimoto, S., S. Kiritani and S. Saito (1980); Perceptual phoneme confusion of frequency compressed speech, Ann. Bull. RILP, 14, 195-200.
3. Sekimoto, S. (1982); Perceptual Normalization of Frequency Scale, Ann. Bull. RILP, 16, 95-101.
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.

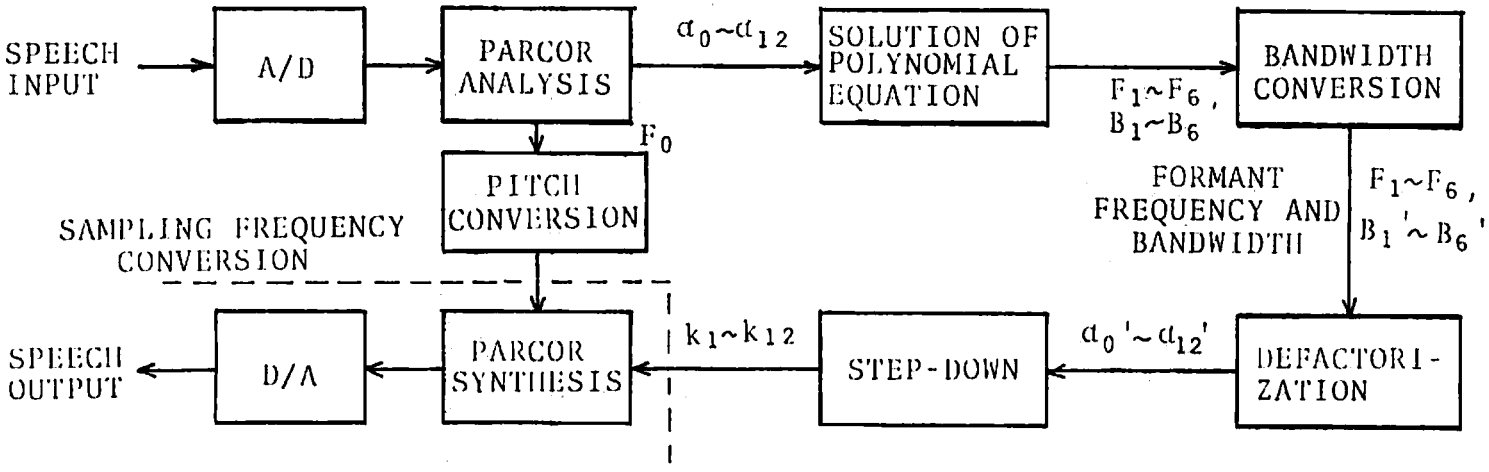

Fig. 1 *Block diagram of the formant peak emphasis processing based on the PARCOR speech analysis-synthesis method.*

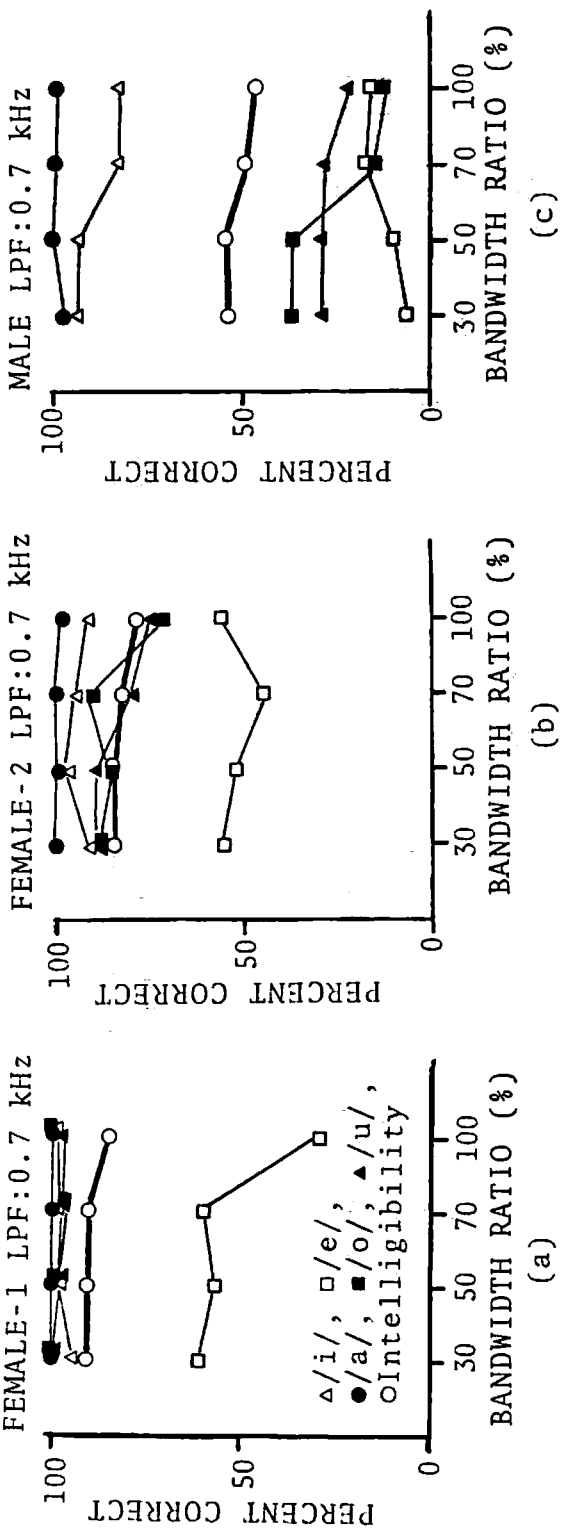

Fig. 2 *Vowel intelligibility and identification rates with a 0.7 kHz low-pass condition: effect of the formant-bandwidth ratio.*

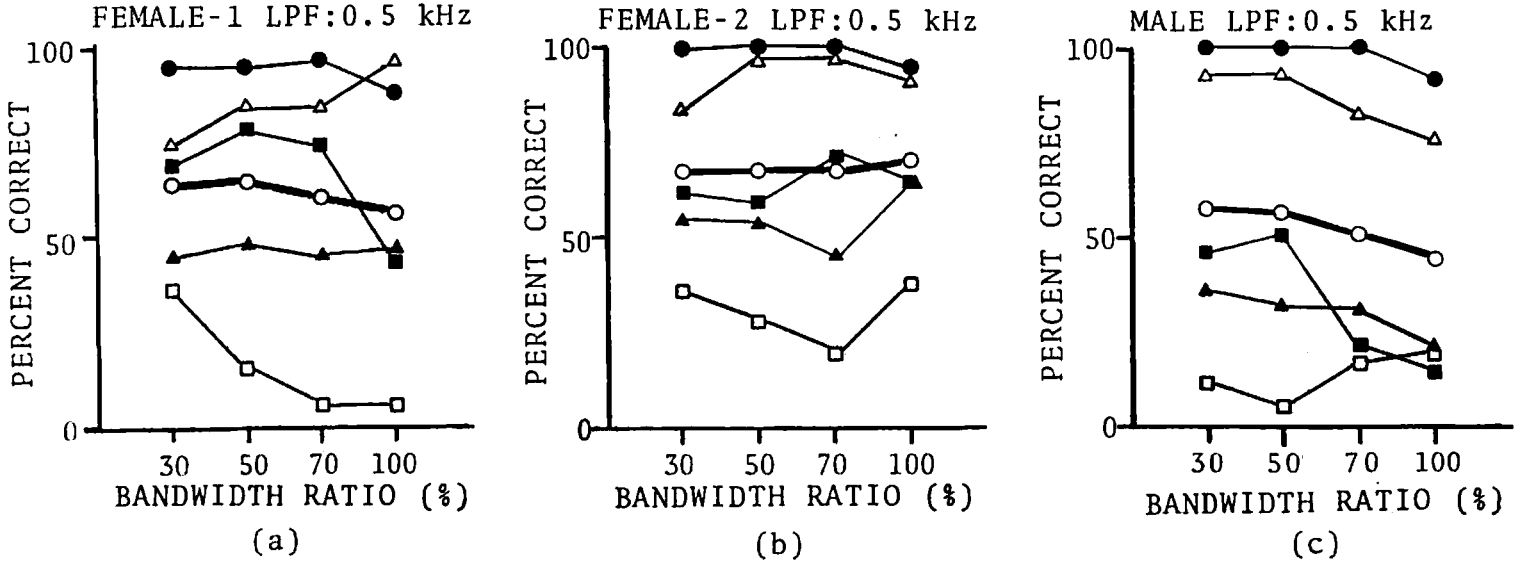

Fig. 3 *Vowel intelligibility and identification rates with a 0.5 kHz low-pass condition: effect of the formant-bandwidth ratio.*