Ann Bull. RILP, No. 14, 65-72 (1980)

NONLINEAR FREQUENCY COMPRESSION SPEECH PROCESSING BASED ON THE PARCOR ANALYSIS-SYNTHESIS TECHNIQUE

Sotaro Sekimoto and Shuzo Saito

1. Introduction

In a previous paper, we proposed a speech processing technique for synthesizing frequency-compressed speech for the hearing impaired[1]. The method is based on the PARCOR analysis-synthesis method. Intelligibility tests of speech processed by this method were carried out with normal hearing persons under lowpass filtering to simulate hearing loss for the high-frequency domain. Detailed results of these experiments are reported elsewhere in this issue[2]. A summary of the results is as follows:

1) Intelligibility was markedly improved for female vowels after frequency compression.
2) Improvement in intelligivility was insignificant for consonants in the case of both male and female speakers.
3) For male vowels, no improvement in the intelligibility score was observed.

Reasons for the unsatisfactory improvements for the consonants and male vowels are considered as follows. In the case of consonants, the high-frequency component which bears the acoustic cues of consonants is lost by the lowpass filtering simulating hearing loss. For instance, when the cutoff frequency of lowpass filtering is 1kHz and the frequency compression ratio is 50%, the frequency component above 2kHz is lost. In such a situation, the acoustic cues in the high-frequency range, e. g., those for palatal stops and palatalized fricatives, will be lost. In the case of male vowels, the low-frequency component is moved into extremely low-frequency range by the frequency compression. Thus, intelligibility of consonants may be improved by adopting higher rate of frequency shift for the high-frequency component of the consonants. Intelligibility of male vowels may be also improved by adopting lower rate of frequency shift on the low-frequency component of vowels.

The present paper thus proposes a new frequency compression technique to compress the frequency scale nonlinearly.

2. Nonlinear Frequency Compression Technique

As a nonlinear frequency compression technique, we previously examined a method to move the pole frequencies downward[1]. However, the overall slope of the spectrum envelope obtained was steeper than that of the original spectrum. To avoid this effect, an extra pole pair was added at the highest end of the frequency range. The spectrum envelopes with and without the extra pole are compared in Figure 1.

This compression method seems to be effective when the degree of compression is weak. However, this method does not yield a satisfactory compensation when the compression ratio is less than 70%. Thus, a new nonlinear frequency compression technique has been proposed in which these problems are avoided.

The new nonlinear frequency compression technique consists of two processes. The first is a nonlinear transformation of the frequency scale. The second process is a linear compression of the frequency axis. Through this process, the high-frequency component of the original speech is moved further downward. Figure 2 shows the calculation method for these processes. In the first process, the digitized speech wave is subjected to preliminary PARCOR analysis in order to extract linear predictor coefficients, pitch period, excitation power and V/UV decision[3]. The spectrum envelope is calculated by the discrete Fourier transform (DFT) from the linear predictor coefficients. Then, the modified spectrum envelope $S'(f)$ is derived from the original spectrum envelope $S(f)$ as

$$S'(f) = S(f')$$

using frequency transformation function $f' = G(f)$.

The following restrictions are required for the nonlinear transformation function:

1) Both lowest and highest frequency of the original bandwidth must be unchanged.
2) The function increases monotonously.

From the modified spectrum $S'(f)$, the autocorrelation function is calculated by the inverse DFT and subjected to final PARCOR analysis in order to extract the PARCOR k-coefficients.

In the second process, the linear frequency compression is performed by simply reducing the sampling frequency of the synthesizer. In this process, the values of the PARCOR k-coefficients are used without modifying the values extracted by final PARCOR analysis.

3. Evaluation

In order to evaluate the efficiency of nonlinear frequency compression, speech compressed by the present technique is synthesized for the Japanese monosyllables. It should be noted here that the cases for the consonants and male vowels are contradictory. While the high-frequency component should be moved into the lower frequency range for the former case, the lower frequency component should be moved upward for the latter case. Therefore, in this study, different frequency transformations were applied to these two cases.

The original speech is lowpass-filtered at 5kHz and sampled at 10kHz. The speech signal is digitized to 10 bits. Preliminary PARCOR analysis is carried out to compute a set of $p = 12$ linear predictor coefficients from the 20 msec of Hamming windowed speech, and analysis is repeated successively every 5 msec. The spectrum envelope is calculated from the linear predictor coefficients by the 256-point FFT. As a frequency transformation function, the same formulation is used for the two cases - consonants and male vowels - by altering the function parameter. As a function which satisfies the restrictions described above, the following function is utilized:

$$G(f) = (1 + r \cdot \frac{1}{B(p,q)} (\frac{f}{f_N})^{p-1} (1 - \frac{f}{f_N})^{q-1}) \cdot f$$

where f is the frequency of the original spectrum component, f_N is one-half the Nyquist frequency,

B(p, q) is the Beta function defined as

$$B(p,q) = \int_0^1 x^{p-1}(1-x)^{q-1}dx,$$

p and q are the parameters which determine the degree of asymmetry and acuteness of the function, and r is the weighting factor of the nonlinearity. The value of r is chosen so that the function G(f) satisfies the restriction of the monotonous increase.

The values of the parameters are chosen as $p = 7$, $q = 3$ and $r = -0.05$ for consonants, and $p = 2$, $q = 20$ and $r = 0.1$ for male vowels. The relationships between the input frequency and the transformed frequency are shown in Figures 3 and 4, respectively, for the two cases. The curves of the nonlinear weighting function are also shown in the same figures. The autocorrelation function is then calculated by computing the 256-point inverse FFT. A new set of PARCOR coefficients is obtained every 5 msec by the final PARCOR analysis from the first 13 terms of the autocorrelation function. The order of the analysis is 12.

The frequency-compressed speech is synthesized for five compression ratios: 100%, 80%, 60%, 50% and 40%. Here, the compression ratio is defined as the percentage of the bandwidth of the compressed speech vs. that of the original one. The sampling frequencies of the synthesis filter are 10kHz, 8kHz, 6kHz, 5kHz and 4kHz, respectively, for the five compression conditions. The output speech signals from the synthesis filter are DA-converted and lowpass-filtered at 5kHz, 4kHz, 3kHz, 2.5kHz and 2kHz, respectively. The resultant bandwidths of the compressed speech have the same values as those of the output lowpass filters.

For the first case, where the frequency component at the higher frequency range is moved downward, analysis and synthesis were performed for Japanese palatal stops and palatal fricatives. Figure 5 shows the spectrum envelope of the vowel portion of /ka/ processed by both the linear and the nonlinear frequency compression techniques. It can be seen that the higher frequency components are moved downward.

For the second case, where the lower frequency component is moved upward for male vowels, the effect of the nonlinear processing was examined for the five Japanese male vowels. Comparisons of the spectrum envelopes processed by both the linear and the nonlinear processing techniques are shown in Figure 6, for vowel /i/ and /o/. It is obviously that the lower part of the spectrum is moved upward.

Intelligibility tests to evaluate the nonlinear processing technique for the two classes of sounds, i.e., consonants which have an acoustic cue in the higher frequency range and male vowels which have very low-frequency components, are now being planned.

Acknowledgment

This study was supported in part by a Grant-in-Aid for Scientific Research (No. 579013) from the Japanese Ministry of Education, Science and Culture.

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, 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
3) Itakura, F. and S. Saito (1969): "Speech Analysis-Synthesis Based on the Partial Autocorrelation Coefficients," Acous. Soc. of Japan Meeting. 2-2-6, 199-120 (in Japanese).
4) Itahashi, S and S. Yokoyama (1974): "A Method of Formant Extraction with Logarithmic Frequency Axis," Report of the Autumn Meeting of the Acoustical Soc. Japan, 235-236 (in Japanese).

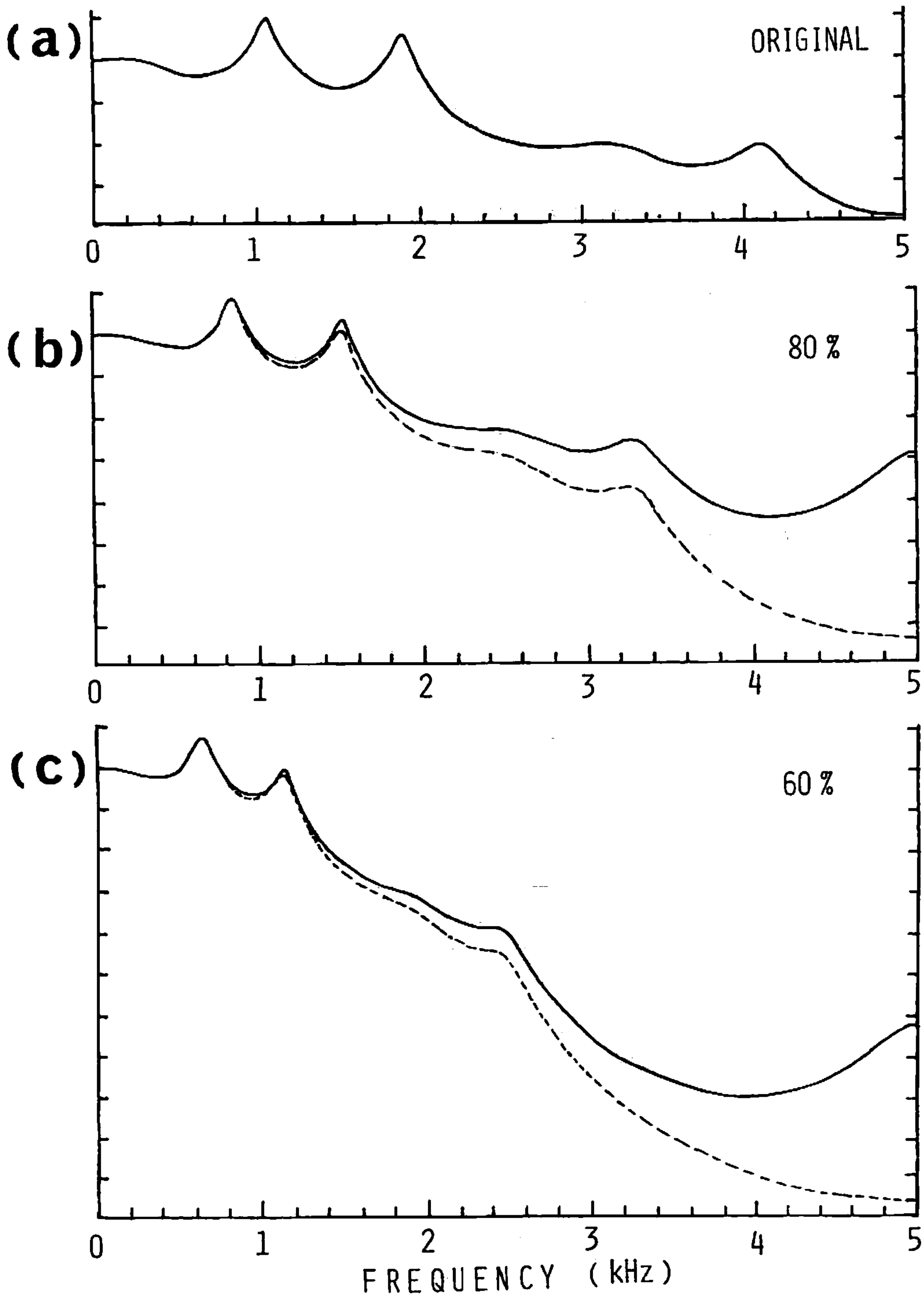

Fig. 1 The spectrum envelopes of the frequency-compressed speech /a/ which is made by lowering the pole frequencies: (a) original frequencies, (b) 80% of the original, (c) 60% of the original. Solid lines show the spectrum envelopes when the extra pole is added.

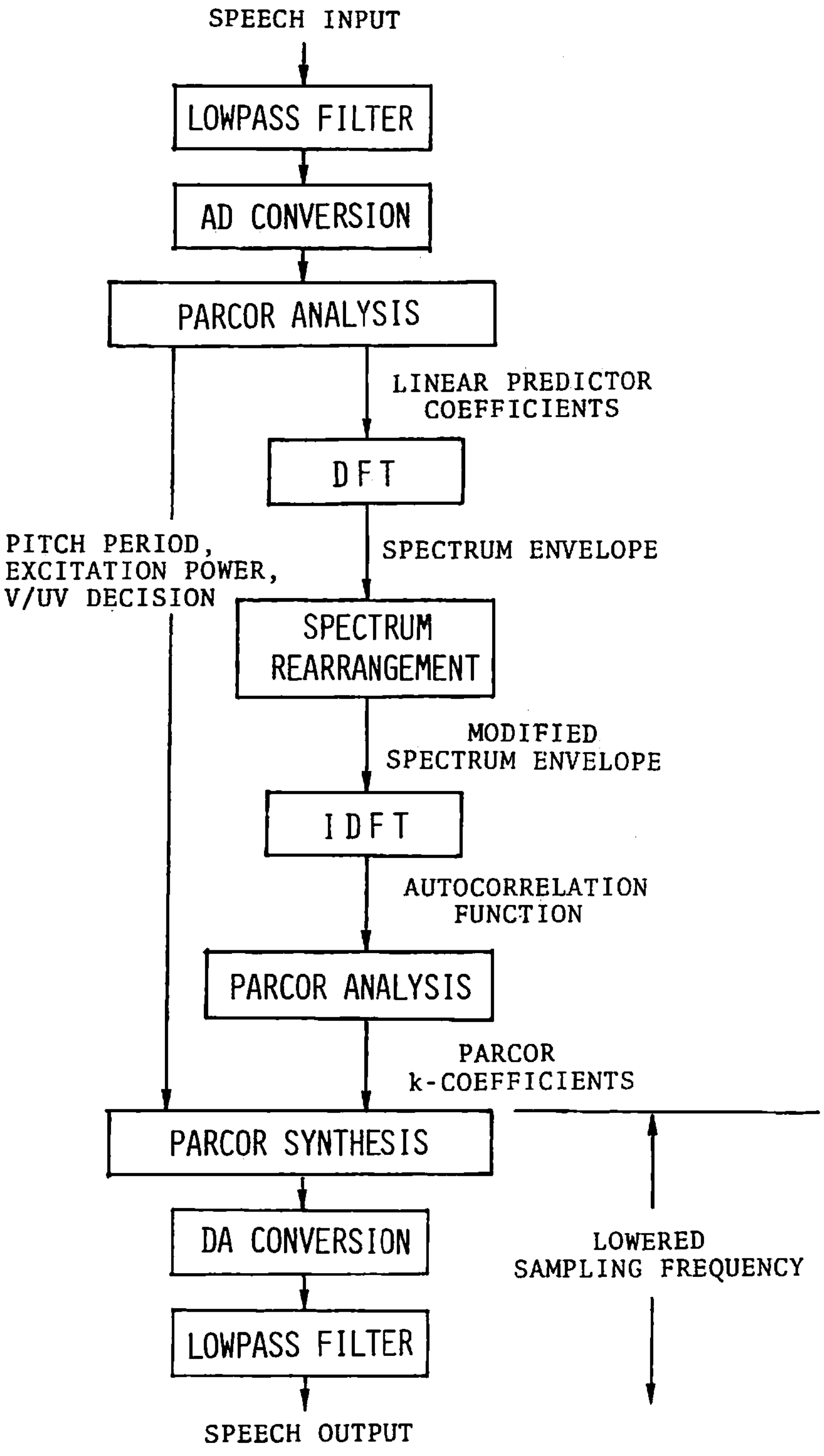

Fig. 2 Calculation process of the nonlinear frequency compression technique.

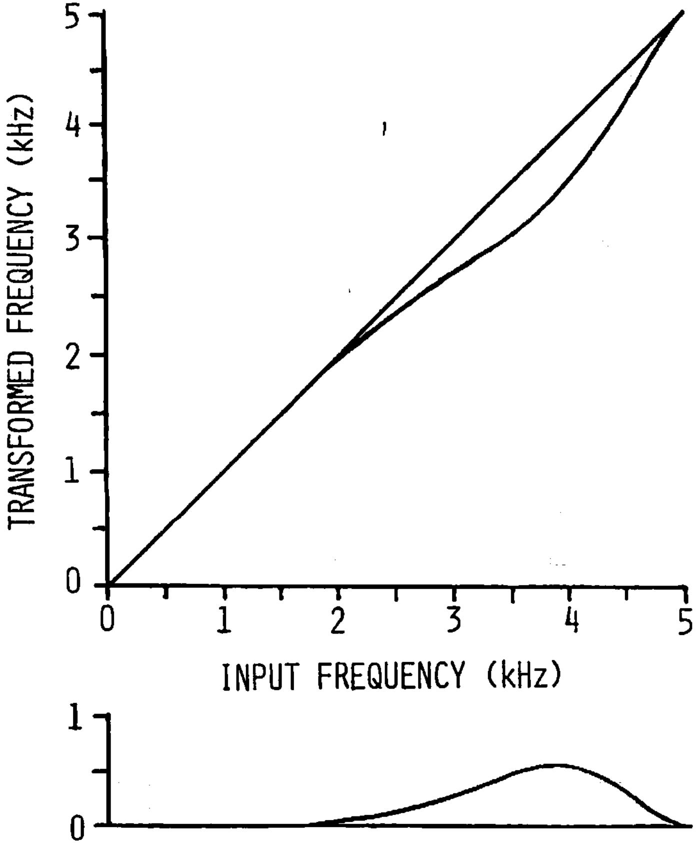

Fig. 3 Nonlinear relationships between the input and transformed frequency used for the male vowels. The nonlinear weighting curve is also shown.

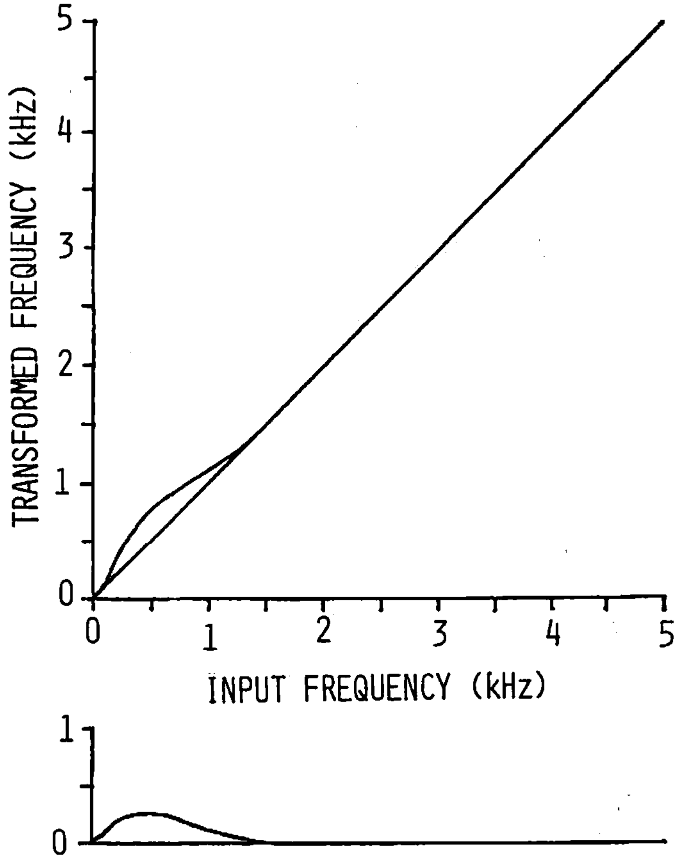

Fig. 4 Nonlinear relationships between the input and transformed frequency used for the male vowels. The nonlinear weighting curve is also shown.

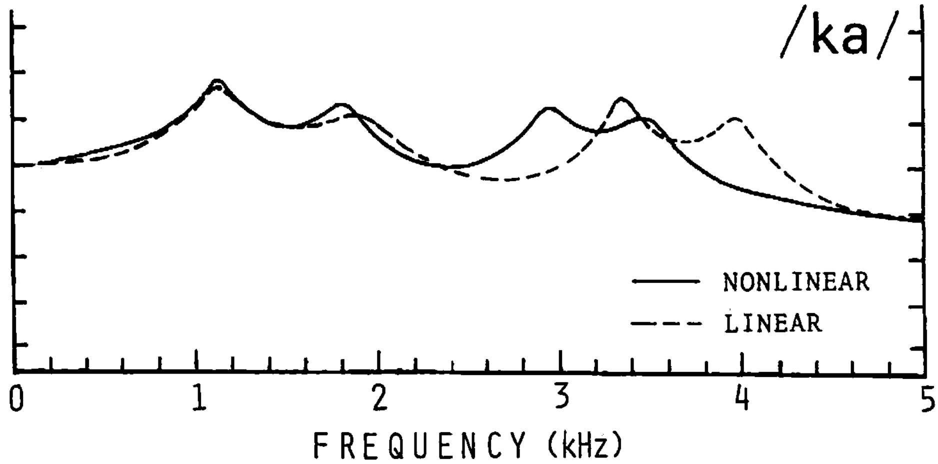

Fig. 5 The spectrum envelope of the vowel portion of /ka/. Comparison is made within the linear and the nonlinear compressions.

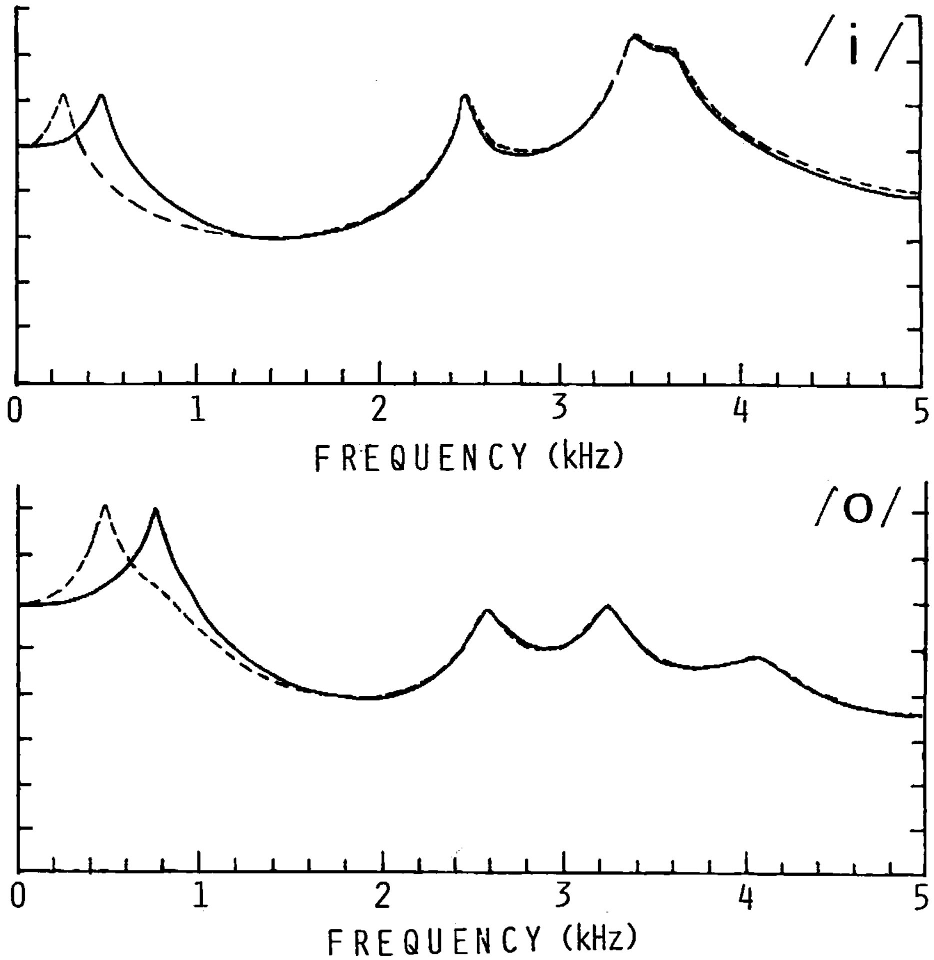

Fig. 6 Comparisons of the linear and the nonlinear compressed spectrum envelopes of the vowels /i/ and /o/.