Ann. Bull. RILP (1983) No. 17, 115-121

INTELLIGIBILITY OF NONLINEAR FREQUENCY COMPRESSED VOWELS

Sotaro Sekimoto

1. Introduction

A frequency compression method for providing intelligible speech for the hearing impaired has been studied by the present authors.[1] Intelligibility tests have been conducted on the processed speech to assess the efficiency of the method quantitatively. Tests were carried out using normal hearing subjects under a lowpass filtering condition which simulated the hearing loss of a deaf person. The results showed that the intelligibility was markedly improved for female vowels after frequency compression. For male vowels. however, the improvement in the intelligibility score was not significant. It can be concluded that this unsatisfactory improvement was caused by an irregular overlap of the fundamental frequency and the first formant frequency. In the case where a vowel has a low first formant frequency, as in the vowel /u/, the significant frequency components around the first formant are moved into the fundamental frequency region by the frequency compression. In order to solve this problem, a nonlinear frequency compression technique has been developed by the authors.[2] By this method, higher frequency components are compressed uniformly, while lower frequency components around the first formant are nonlinearly less compressed.

In the present study, the intelligibility of nonlinear frequency compressed vowels uttered by a male speaker was measured using normal hearing subjects with a simulated hearing loss.

2. The Nonlinear Frequency Compression Method

The calculation process for the nonlinear frequency compression is schematized in Fig. 1. The speech signal was digitized and then subjected to a preliminary PARCOR analysis to extract the PARCOR coefficients, pitch period, excitation power and voiced/unvoiced distinction. The PARCOR coefficients were then used to calculate the spectrum envelope. Other parameters were used to control the excitation source of the synthesizer at later stages. The calculation of the spectrum envelope was done by a discrete Fourier Transform (DFT). The spectrum $S(f)$ was transformed into the modified spectrum $S'(f)$ as

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

using the transformation function $f' = G(f)$, where the following restrictions are imposed:

1) Both the lowest and the highest end in the frequency range are unchanged.
2) The function is a monotone increasing function.

From the modified spectrum $S'(f)$, the autocorrelation function was calculated using the inverse DFT. Then, the final PARCOR analysis was made to extract the new PARCOR coefficients. These PARCOR coefficients were then sent to the

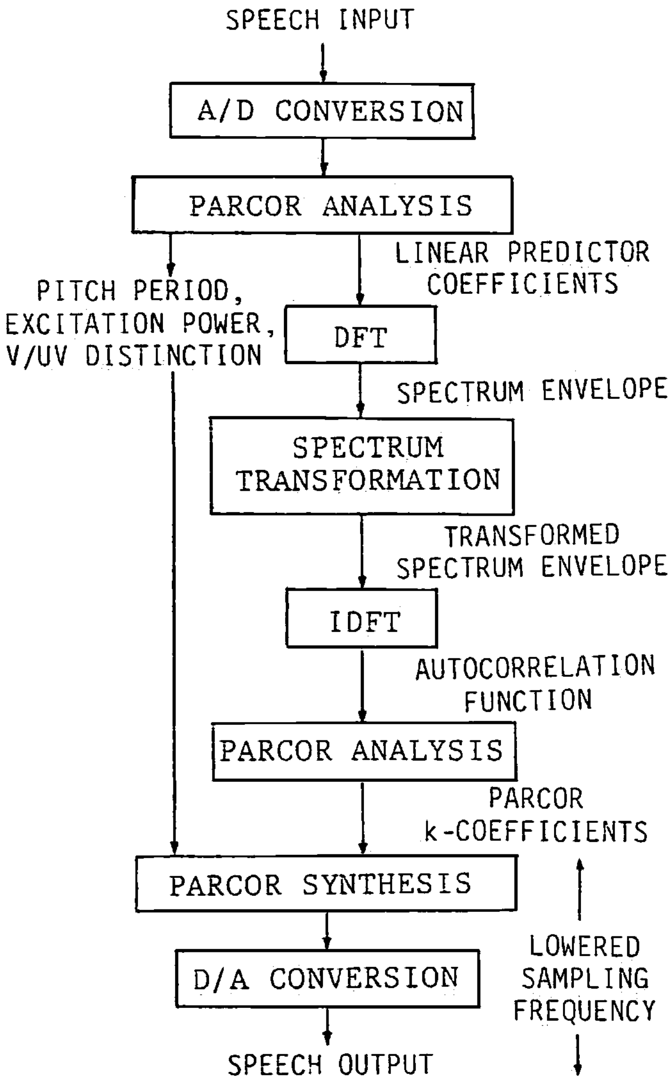

Fig. 1 *Calculation process for the nonlinear frequency compression method.*

synthesizer to generate the frequency compressed speech. The frequency compression was made by simply reducing the sampling frequency from that of the analysis stage. Thus, the nonlinear frequency compression was performed as a combination of nonlinear frequency transformation and linear frequency compression.

3. Method

3.1 Measurement

The speech materials were five steady-state Japanese vowels uttered by a male speaker. The speech samples processed by the method described above were subjected to an intelligibility test. The tests were carried out in a soundproof room for every combination of the experimental conditions. The experimental conditions employed in the intelligibility test were as follows.

(1) Frequency Compression Ratio

The frequency compression ratio was defined as the percent ratio of the sampling frequency of the synthesizer to that of the analyzer in the speech processing described above. The ratios 100%, 80% and 60% were examined. The fundamental frequency was maintained in its original state even when the frequency was compressed.

(2) Cutoff frequency of the lowpass filter

A lowpass filter was used to simulate the frequency response of the hearing impaired. The cutoff frequency was 0.5 kHz. The slope of the filter was more than 120 dB per octave.

The speech stimuli were presented in a random order to one ear using headphones at a level of 10 dB OTR (Ortho-telephonic response). One male and three female subjects were asked to identify each stimulus as one of the five Japanese vowels.

3.2 Processing of the Speech Material

The speech materials were recorded in a soundproof room. The processing of the speech was performed on a PDP-11/34 computer. The speech signal was lowpass-filtered at 5 kHz; sampled at 10 kHz; and digitized to 10 bits. A preliminary PARCOR analysis was carried out to extract the PARCOR coefficients. The analysis was made every 5 msec on 30 msec of Hamming-windowed speech. The order of the analysis was 12. The pitch period was determined by detecting the autocorrelation peak of the residual wave, with the spectral deformation caused by the vocal tract resonance removed. The spectrum envelope was calculated by the 256-point FFT. As the nonlinear transformation function, the following function was used.

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

where f was the frequency of the original spectrum component, and f_N was the Nyquist frequency.

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

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

where p and q were the parameters which determined the degree of asymmetry and the acuteness of the function, and r was the weighting factor of the nonlinearity. The value of r was chosen so that the function G(f) satisfied the restriction of the monotonous increase. The value of the parameters were chosen as $p = 2$, $q = 31$ and $r = 1$. The relationships between the input frequency and the transformed frequency are shown in Fig. 2, and an example of the transformed spectrum envelope compared to that of the original is shown in Fig. 3. The autocorrelation function was then calculated by the 256-point inverse FFT. A new set of PARCOR coefficients was obtained every 5 msec by the final PARCOR analysis. The order of the analysis was 12. The sampling frequencies of the PARCOR synthesizer were 10 kHz, 8 kHz and 6 kHz, respectively, for the three compression ratios. The output speech signals from the synthesizer were DA-converted and lowpass-filtered at 5 kHz, 4 kHz and 3 kHz, respectively.

4. Result

Fig. 4 shows the relation between the vowel intelligibility and the frequency compression ratio. In the figure, the results are compared for nonlinear compressed vowels and linear compressed ones. The solid line shows the result for nonlinear compressed vowels, and the broken line shows that for linear compressed ones. The intelligibility score was greater for the linear compressed vowels when the frequency compression ratio was 100% and 80%. When the frequency compression ratio was 60%, on the other hand, the intelligibility score was greater for the nonlinear compressed vowels.

Fig. 5 shows the articulation scores for the respective vowels. The compression conditions were compared for the three cases where the frequency was not compressed; was linearly compressed to 60%; and was nonlinearly compressed to 60%. When the frequency was not compressed, the articulation scores for /o/ and /a/ were nearly zero. By the linear frequency compression, the articulation scores for these vowels were improved, but for the vowels /u/ and /e/, the articulation scores were reduced. When the frequency compression was made nonlinearly, the deterioration in the articulation score for /u/ was markedly improved. For the vowels /a/, /e/ and /i/, on the other hand, the articulation scores were reduced.

5. Remarks

In this study, intelligibility scores were measured for linear and nonlinear compressed male vowels under a lowpass condition which simulated a hearing impairment. It was observed that nonlinear frequency compression was effective for improving the deterioration in the articulation score of the vowel /u/ caused by ordinary linear frequency compression. On the other hand, nonlinear frequency compression was rather injurious to the other vowels except /o/. Thus, the nonlinear transformation function should be determined so that vowel intelligibility is maxi-

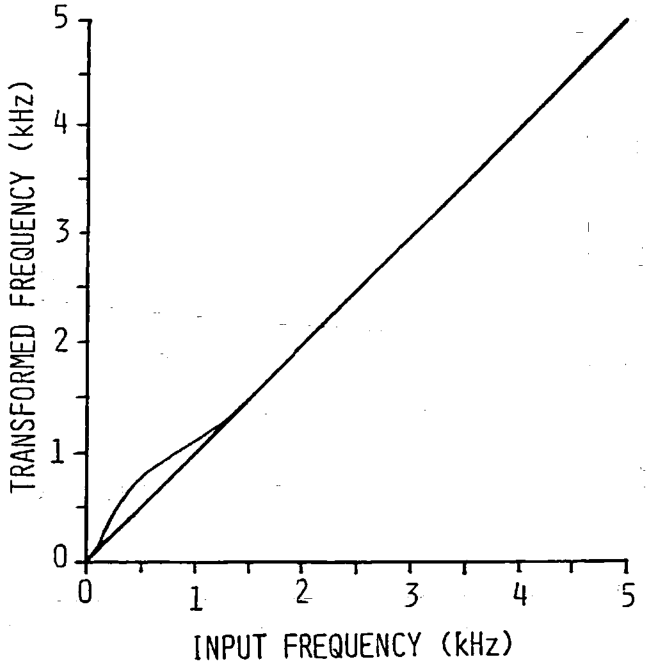

Fig. 2 *Nonlinear relationships between the input and the transformed frequency.*

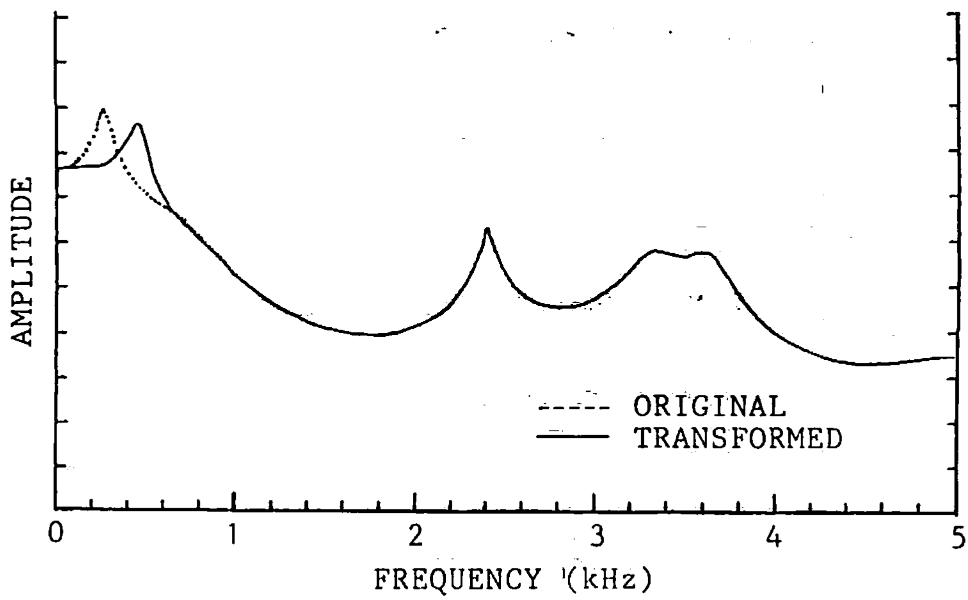

Fig. 3 *Comparison of the original and the transformed spectrum envelopes for the vowel /u/.*

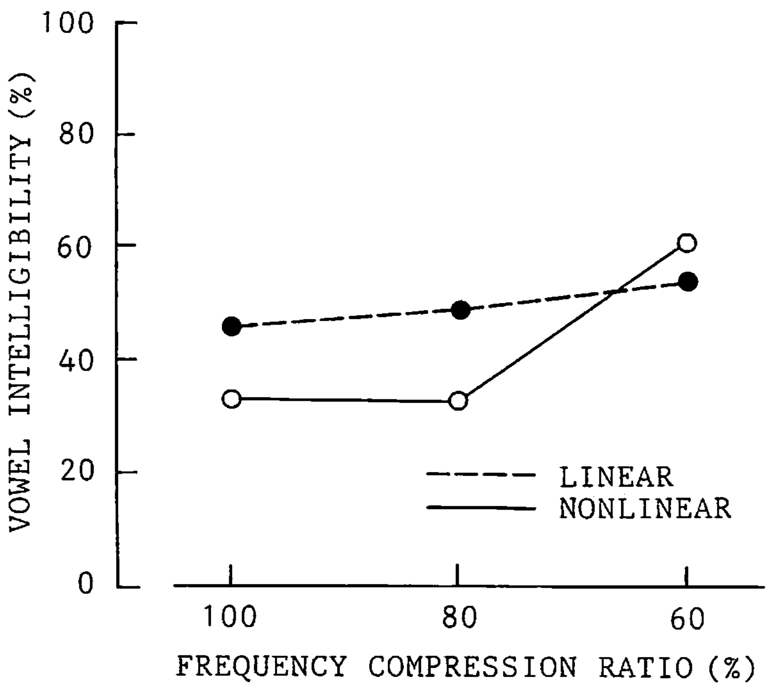

Fig. 4 *Vowel intelligibilities for linear and nonlinear compressed male vowels. The cutoff frequency of the lowpass filter which simulated hearing impairment was set at 0.5 kHz.*

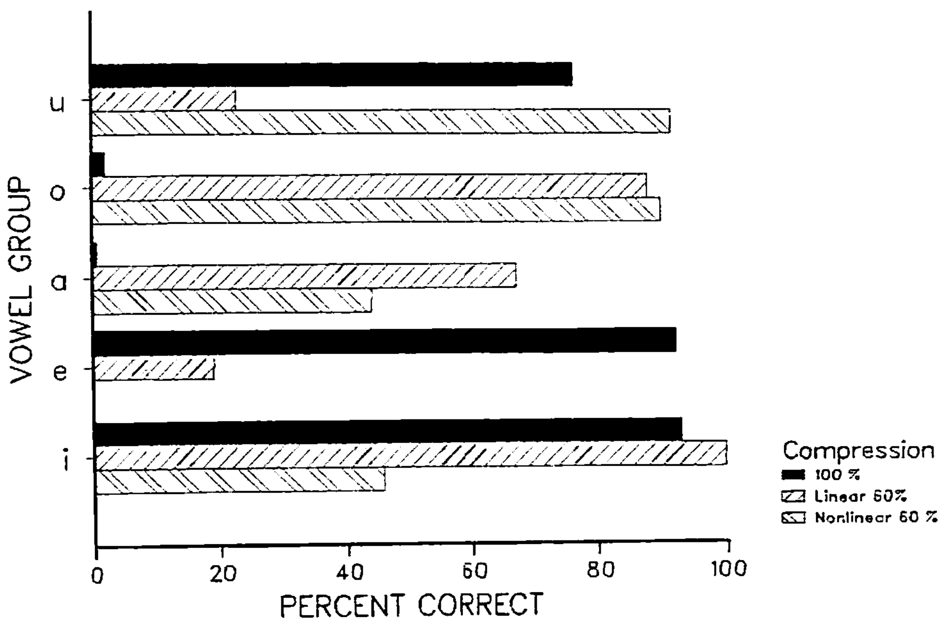

Fig. 5 *Articulation scores for the respective vowels. The results were compared for three compression conditions: non-compressed, linear 60% compression and nonlinear 60% compression.*

mized. How to determine the characteristics of a nonlinear mapping function which would satisfy such a requirement remains to future research.

Acknowledgement

The author wishes to thank Professor Shuzo Saito for suggestions.

This work was supported in part by a Grant-in-Aid for Scientific Research (No. 57780024, No. 57880009) 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 lowpass filtered condition. Ann. Bull. RILP, 14, 181-193.
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. 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).