Ann.Bull.RILP(1984)No.18, 119-125

INTELLIGIBILITY OF NONLINEAR, FREQUENCY COMPRESSED SPEECH

Sotaro Sekimoto

1. Introduction

A frequency compression technique to provide intelligible speech for the hearing impaired has been studied by the present authors.[1] Intelligibility tests have been made to assess the efficiency of the method quantitatively. Tests were carried out using normal hearing subjects with a hearing loss simulating that of the hearing impaired. The results showed that the intelligibility was markedly improved for female vowels (hereafter vowels uttered by female speakers will be termed "female vowels", and so forth) after frequency compression. For male vowels and for male and female consonants, however, the improvement in the intelligibility score was not significant. If these intelligibility scores could be increased, the overall scores would be improved. Among these, male vowels with their original F_0 and male consonants showed less deterioration, and especially, male consonants sometimes showed an improvement, if the cutoff frequency of the low-pass filter was set within a range of from 1.0 to 1.5 kHz.

In the case of male vowels, the fundamental frequency was maintained at its original value in the input speech, because the lowering of the fundamental frequency in proportion to the frequency spectrum compression made the pitch sensation unnaturally low. If a vowel has a low first formant frequency, as in the vowel /u/, the first formant F_1 approaches the F_0 as a result of the frequency compression, and an irregular spectral peak appears. This problem can be avoided if F_0 and F_1 can be somewhat kept separate by moving either the F_0 downward or the F_1 upward even when the frequency compression is made. For the former case in which the F_0 moved downward, a processing technique was developed and its efficiency was tested for female vowels with an intelligibility test using processed speech samples.[1] For male vowels, although not yet tested, a similar efficiency can be expected. For the latter case in which the F_1 is moved upward, a nonlinear frequency compression technique has been developed by the authors and reported previously in this Bulletin.[2] By this method, higher frequency components are compressed uniformly, while lower frequency components around the first formant are less compressed nonlinearly.

In the case of consonants, the intelligibility of consonants will be improved if the degree of the compression is increased because the frequency component of consonants which contains the principal cue to their characterization comes within the passband of the lowpass filter simulating the hearing loss. Therefore, the greater degree of the frequency compression will bring the greater improvement in the intelligibility score. For vowels, however, the optimum degree of the compression exists [1], and the degree will be less than that for consonants. Moreover, the linear frequency compression is desirable for vowels in the frequency range below the F_3 in order to maintain their original phonetic quality. Thus, the frequency compression scheme is desirable, in which frequency components are compressed uniformly within the frequency range between F_1 and F_2, and nonlinearly stronger above F_2. The nonlinear frequency compression technique developed by the present authors can be adapted to such a frequency compression scheme.[3)]

In the present study, the intelligibility of nonlinear frequency compressed speech was measured to evaluate the efficiency of the nonlinear frequency compression technique improving consonant identification.

2. The Nonlinear Frequency Compression Method

The nonlinear frequency compression was accomplished through two processes, as shown in Fig. 1, a nonlinear frequency transformation and a linear frequency compression. The calculation process for the nonlinear frequency compression is schematized in Fig. 2.

First, the nonlinear frequency transformation was made. The speech signal was digitized and then subjected to a preliminary PARCOR analysis to extract the linear predictor coefficients. The linear predictor coefficients were then used to calculate the frequency spectrum envelope and the residual signal. The frequency spectrum envelope was calculated by a discrete Fourier Transform (DFT) of the linear predictor coefficients. 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 were imposed in order not to reduce the number of poles.

1) Both the lowest and the highest points in the frequency range were left unchanged.

2) The function was a monotone increasing function.

From the modified spectrum S'(f), an autocorrelation function was calculated using the inverse DFT. Then, the final PARCOR analysis was made to extract the PARCOR coefficients. The PARCOR coefficients were sent to the PARCOR synthesizer to generate the frequency compressed speech. The synthesizer was driven by the residual signal. The residual signal was given as an output of the inverse filter whose coefficients were the linear predictor coefficients derived during the preliminary PARCOR analysis.

The second process, a linear frequency compression, was made by simply reducing the sampling frequency of the synthesizer at the analysis stage. Thus, the nonlinear frequency compression was performed as a combination of the first nonlinear frequency transformation and the second linear frequency compression.

3. Method

3.1 Measurement

The speech materials were 100 Japanese monosyllables uttered separately by a male speaker. The processed speech samples 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. The ratios (i) 100% and (ii) 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 frequencies were (i) 1.5 kHz and (ii) 0.7 kHz. The slope of the filter was -42 dB per octave.

(3) Nonlinear Curve of the Transformation Function

As the nonlinear transformation function, the following function was used

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

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 following five parameter sets of p, q and r were examined.

(L) r=0 : linearly compressed
(A) p=5, q=3 and r=-5
(B) p=5, q=2 and r=-5
(C) p=5, q=3 and r=-3
(D) p=5, q=2 and r=-3

The relationships between the input frequency and the transformed frequency for these parameters are shown in Fig. 3.

The speech stimuli were presented in a random order to one ear using headphones at a level of 10 dB OTR (ortho-telephonic response). A male subject was asked to identify each stimulus as one of 100 Japanese monosyllables.

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 VAX-11/780 computer. The speech signal was lowpass-filtered at 10 kHz; sampled at 20 kHz; and digitized to 12 bits. A preliminary PARCOR analysis was made every 5 msec on 30 msec of Hamming-windowed speech. The order of the analysis was 24. The spectrum envelope was calculated by a 1024-point FFT. The autocorrelation function was calculated with the 1024-point inverse FFT. A new set of PARCOR coefficients was obtained every 5 msec by the final PARCOR analysis. The order of the synthesizer filter was 24. The sampling frequencies of the PARCOR synthesizer were 20kHz and 12kHz, respectively, for the two compression ratios. The output speech signal from the synthesizer was DA-converted and lowpass-filtered.

4. Results

The intelligibility scores are shown in Figs. 4 and 5 for the lowpass filter cutoff frequencies of , 1.5 kHz and 0.7 kHz, respectively. In these figures, the results for each nonlinear curve are compared. The frequency compression ratios are 60%. The solid line shows the intelligibility for vowels, and the broken line shows that for consonants.

When the cutoff frequency of the lowpass filter was 1.5 kHz, the intelligibility scores for the consonants varied with respect to the nonlinear curve and reached maximum with the curve B. The intelligibility scores for the vowels were almost 100%. When the cutoff frequency of the lowpass filter was 0.7 kHz, on the other hand, the intelligibility for the consonants was nearly constant regardless of the nonlinear curve. The intelligibilities for vowels were not affected by the shape of the nonlinear curve.

5. Remarks

In this study, intelligibility scores were measured for linear and nonlinear, compressed male speech under a lowpass condition which simulated the frequency characteristics of a hearing impairment. It was observed that the intelligibilities for consonants were affected with regard to the shape of the nonlinear curve when the cutoff frequency of the lowpass filter was 1.5

kHz, but was not affected when the cutoff frequency of the lowpass filter was 0.7 kHz. This result suggests that nonlinear compression is successful in raising the intelligibility of consonants because the nonlinear portion of the transformation curve is included within the passband of the filter when the cutoff frequency is 1.5kHz but is not included when it is 0.7 kHz. For vowels, on the other hand, the intelligibility was not affected by the shape of the nonlinear curve. This result shows that the expectation that nonlinear compression would not affect the vowel identification was confirmed. The problem of how to determine the turning point in the transformation function from linear to nonlinear has been left unsolved.

Acknowledgement

This work was supported in part by a Grant-in-Aid for Scientific Research (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. (1983); Intelligibility of nonlinear frequency compressed vowels, Ann. Bull. RILP, 17, 115-121.
3) Sekimoto, S. and S. Saito (1980); Nonlinear frequency compression speech processing based on the PARCOR analysis-synthesis technique, Ann. Bull. RILP, 14, 65-72.

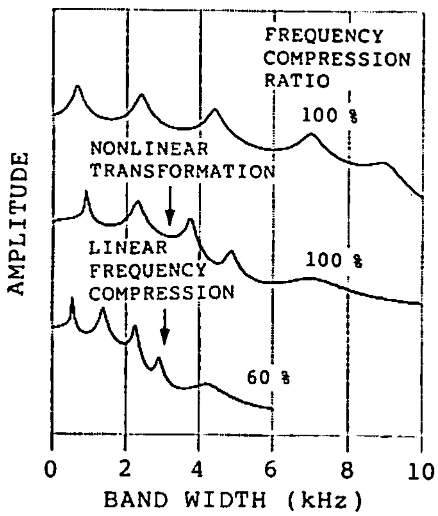

Fig. 1. The principle of nonlinear frequency compression. The nonlinear frequency compression is accomplished using two processes: a nonlinear frequency transformation and a linear frequency compression.

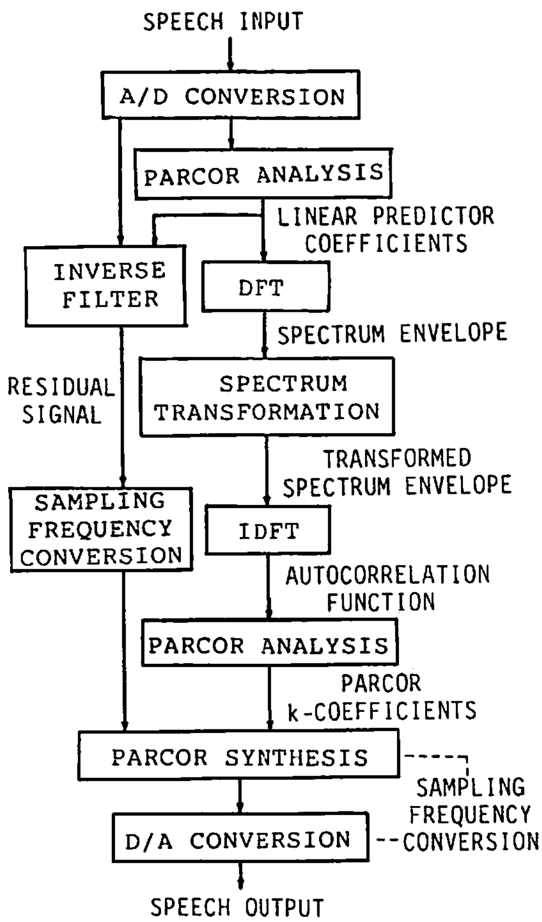

Fig. 2. The calculation process for the nonlinear frequency compression method.

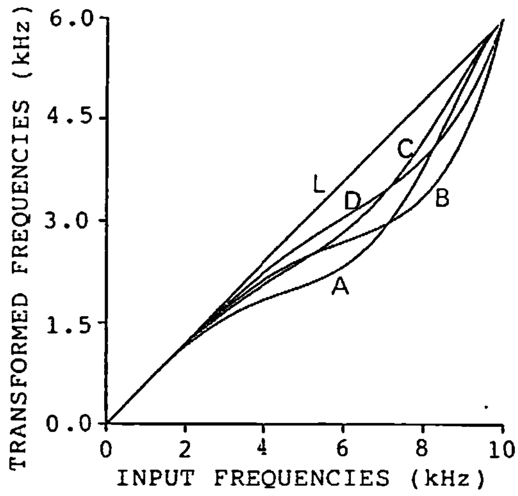

Fig. 3. Relationships between the input and the transformed frequencies for several frequency transformation curves.

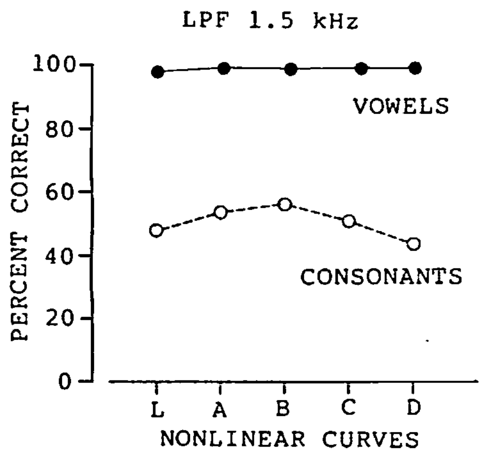

Fig. 4. Intelligibilities for the linear and the nonlinear frequency compressed speech when the cutoff frequency of the lowpass filter was 1.5 kHz.

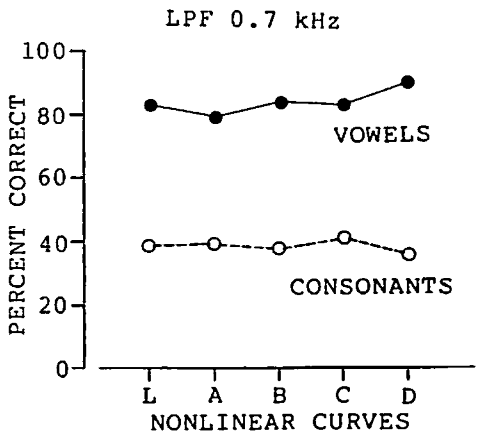

Fig. 5. Intelligibilities for the linear and the nonlinear frequency compressed speech when the cutoff frequency of the lowpass filter was 0.7 kHz.