Ann. Bull. RILP, No. 13, 133-136 (1979)

FREQUENCY COMPRESSION TECHNIQUES OF SPEECH USING LINEAR PREDICTION ANALYSIS-SYNTHESIS SCHEME

Sotaro Sekimoto, Shigeru Kiritani and Shuzo Saito

A large majority of people with sensorineural hearing impairment still have some residual hearing in the low frequency range. With the aim of providing intelligible speech signals to this type of person, several techniques of frequency compression for transposing high frequency information in speech signals into the usable lower frequency range have been proposed and tested.[1] It appears, however, in frequency compression by these methods,there are some additional distortions of the speech spectrum. For the quantitative assessment of frequency compressed speech, a technique which permits more precise manipulation of the spectrum pattern is needed. In the present study, new techniques of frequency compression based on linear prediction analysis-synthesis of speech have been tested.

Principle of the Frequency Compression Technique

The speech signal was subjected to an ordinary PARCOR analysis to extract PARCOR coefficients, pitch period, excitation power and voiced/unvoiced decision.[2] The values of the parameters were modified to transpose the poles of the LPC spectrum into the lower frequency range. The frequency compressed speech was synthesized through a PARCOR synthesis filter using modified parameter values. Two methods for the modification of the parameter values are presently being tested. The computational process is shown in Figure 1.

The first method pertains only to the linear frequency compression of the spectrum. This can be achieved by simply reducing the sampling frequency while using the original values of the PARCOR coefficients without modification in the synthesis process. For example, when the sampling frequency of the original speech is 10 KHz and the desired frequency compression ratio is 0.8, the sampling frequency in the synthesis is reduced to 8 KHz. The number of time samples per frame interval and per pitch pulse interval are adjusted according to the new sampling frequency. By this method, all the pole frequencies, both real and complex, are lowered by a given frequency compression ratio, with the spectrum of the synthesized speech being the linear frequency compression of the original spectrum.

The second method is based on the direct transformation of the pole frequencies. This method enables nonlinear frequency compression of the spectrum. The PARCOR coefficients for the frequency compressed speech were derived in the following way. Based on the values of the linear prediction coefficients obtained for the original speech, complex pole frequencies were calculated by a standard root-solving procedure. These pole frequencies were transposed into the lower frequency range. Any type of functional formula between the original and the new pole frequencies can be adopted for the transformation of the pole frequencies. The linear predictive coefficients corresponding to these new pole frequencies were then calculated, and these were converted into the PARCOR coefficients to be used in the synthesis filter.

In the present study, all the signal processing was performed on a laboratory mini-computer (PDP-11/34). Naturally, the amount of the required signal processing was less for Method 1 than for Method 2. In Method 1, most of the processing time was required for the extraction of pitch information. Because of this, processing time can be reduced considerably if the residual wave, which can be obtained as an output of the PARCOR analysis filter, is used for the excitation of the synthesis filter. In this way, processing time can be made small enough so that the real-time hardware implementation will become possible.

A Preliminary Test on the Intelligibility of the Japanese Monosyllables

As a first step in the evaluation of the feasibility of the present frequency compression methods, the frequency compressed speech by Method 1 was synthesized for the 100 Japanese monosyllables, and an informal listening test on the intelligibility of the test sounds was performed.

The original speech, uttered by an adult female, was lowpass filtered to 5 KHz, sampled at 10 KHz and digitized in 10 bits. PARCOR analysis was made every 5 msec on 20msec of Hamming windowed speech. The order of the analysis was 12 and the pitch period was determined by detecting the maximum peak in the autocorrelation function of the residual wave.[3] The voiced/unvoiced judgement was performed based on the values of the maximum peak in the autocorrelation of the residual wave and the first order PARCOR coefficient, which represents the overall slope of the spectrum.[4] The four conditions of the frequency compression ratio were studied. The compression ratios used were 80%, 70%, 60% and 50%, namely the frequency bands were reduced to 4 KHz, 3.5 KHz, 3 KHz and 2.5 KHz respectively. With regard to the fundamental frequency, two series of the test sounds were synthesized. In one series, the fundamental frequency was kept the same as the original speech and in another series, the fundamental frequency was lowered in the same ratio as the frequency compression ratio so that the spectral density was kept homologous with that of the original speech. The stimuli were presented to the subjects in a random order through headphones in a sound-proof room. Two male subjects with normal hearing participated in the listening test.

The results show that the intelligibility for vowels was nearly 100% for all the conditions of the compression ratio. Naturally, the intelligibility was lower for the consonants. It can be seen that the lowering of the intelligibility was more apparent for the fricatives than for the other types of consonants. It can also be observed that the intelligibility was higher for the test sounds with lowered fundamental frequencies.

A more complete study on the quality of speech produced by the present frequency compression techniques is now in progress.

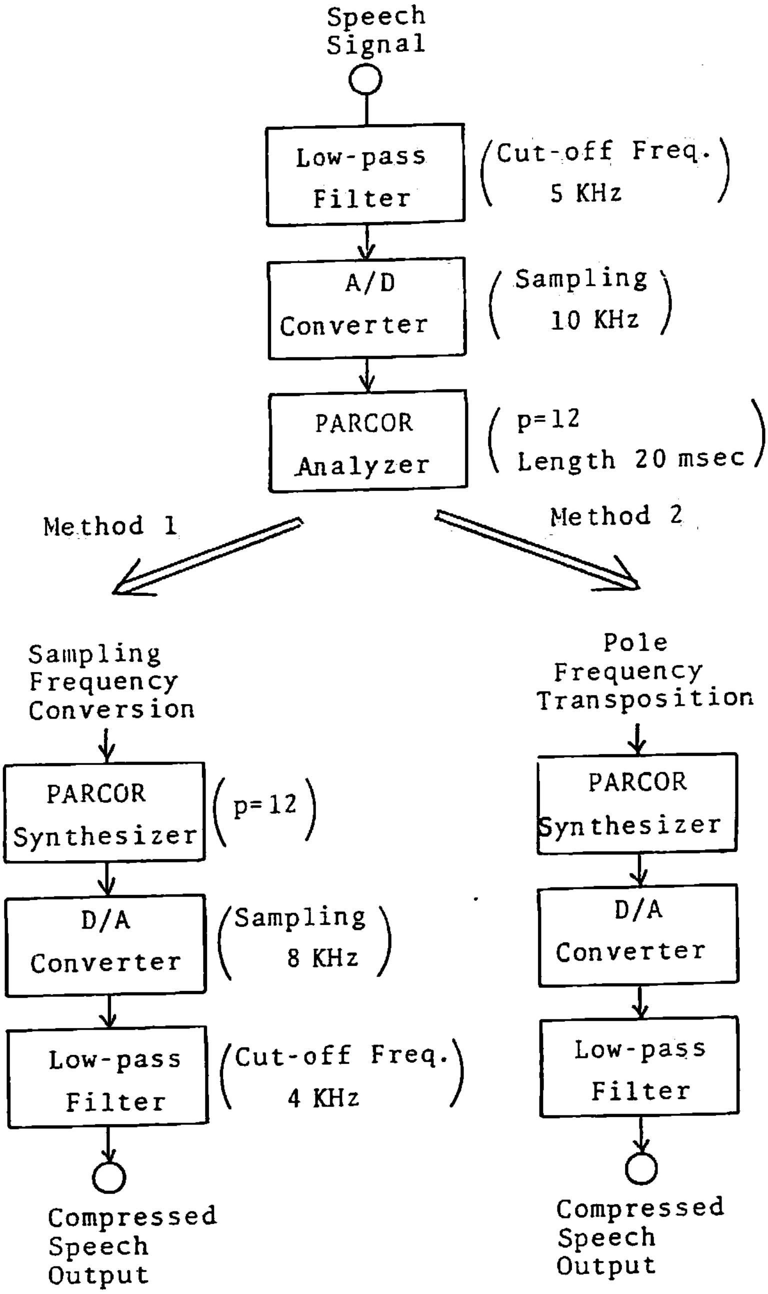

Fig. 1. Schematic diagram of the frequency compression. The parameter values used in the preliminary listening test are shown in parentheses.

References

(1) Mazor, M. (1977); Moderate frequency compression for the moderately hearing impaired. J.A.S.A., Vol. 62, No. 5, 1273-1278.

(2) Itakura, F. and S. Saito (1969); Speech Analysis-Synthesis Based on the Partial Autocorrelation Coefficients. Acous. Soc. of Japan Meeting. 2-2-6, p. 199-120 (In Japanese).

(3) Itakura, F. and S. Saito (1968); Analysis Synthesis Telephony based on the Maximum Likelihood Method. Proc. 6th ICA, Paper C-5-5.

(4) Kitawaki, N., Y. Tohkura and K. Itoh (1977); A Real-time PARCOR Speech Analysis Synthesis System Using a Small Computer. (E. C. L. Tech. Journ.), NTT Japan, Vol. 26, No. 12, 105-117. (In Japanese).

Acknowledgment

This study was supported in part by Grants in Aid for Scientific Research (No. 479014), Japanese Ministry of Education.