RESTORATION OF WIDE BAND SIGNAL FROM TELEPHONE SPEECH USING LINEAR PREDICTION ERROR PROCESSING

Hiroshi Yasukawa

NTT Optical Network Systems Laboratories
Nippon Telegraph and Telephone Corporation
1-2356 Take, Yokosuka-shi, Kanagawa, 238-03 Japan

ABSTRACT

This paper describes a system that can enhance the quality of speech signals that are severely band limited during regular telephone speech transmission. We have already proposed a spectrum widening method that utilizes aliasing in sampling rate conversion and digital filtering for spectrum shaping. This paper proposes a new method using linear prediction (LP) residual error processing and parameter (coefficients) mapping using linear extrapolation that offers improved performance in terms of the spectrum distortion characteristics. Real time hardware implementation procedures are discussed. The proposed method can effectively enhance speech quality.

1. INTRODUCTION

Signal enhancement techniques for speech and audio signals are being widely studied, for example, telephone band signals have been converted into wide band signals [2],- [13]. In telephone communication, if the quality of conventional telephone speech is enhanced for traditional telephone connections, the effectiveness of all telephone communication services will be improved.

ISDN offers a wideband speech signal with a frequency band from 0.05 to 7 kHz according to ITU-TS (old CCITT) recommendation G.722 [1]. However, ISDN customers can not now obtain high quality speech when they call a conventional telephone (0.3 to 3.4 kHz frequency band) user. This is one reason why quality enhancement is necessary.

There are many approaches to enhance speech quality [2],- [9]. We have proposed a very simple method of improving the sound quality of band limited (telephone) speech signals [5], [6], [10]. It is a spectrum widening method that utilizes aliasing effects in sampling rate conversion with digital filtering for spectrum shaping.

Other approaches are based on the assumption that the upper band spectrum envelope can be obtained since the lower and upper band spectrum are correlated. This leads to the adoption of methods such as the statistical recovery function [3], codebook mapping using vector quantization [4], envelope prediction by filtering time trajectories of LPC-cepstral coefficients [8], linear mapping [9], neural net [9], and, in particular, vector quantization [3], [4], [8], [9].

This paper proposes a new method that reduces the spectral distortion and achieves better speech quality than conventional techniques [5], [6].

2. SPECTRUM EXTRAPOLATION

In this paper, we propose a simple method to extrapolate the spectrum envelope from the narrow (telephone) band to the wide band. An appropriate transform between the envelopes of the narrow band and wide (namely upper) band signals can be achieved in several ways [3], [4], [7], [8] as mentioned above.

In practice, we use classification based on the gradient of the spectrum envelope. A linear gradient of the spectrum envelope, namely the slope of the smoothed spectrum envelope, can be computed from LP coefficients. The spectrum envelope of the out band is extrapolated by linear extrapolation. The LP parameters of the wide band signal are thus generated from the narrow band signal.

Fig. 1 shows the linear approximation (dotted line) of the spectrum envelope. From this the spectrum envelope can be classified. As can be seen in Fig.1, the extrapolation of the out-band spectrum is made by extending the slope of the smoothed spectrum envelope of the band limited (in-band) signal.

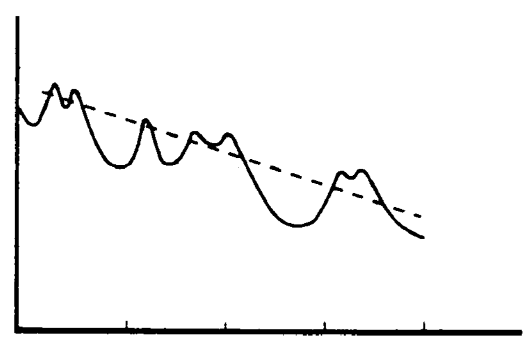

Fig. 1 An example of spectrum envelope and its smoothed slope

In general the spectra of consonants are flat as a white signal. However, in this method, it is assumed that the slope of the spectra of consonants follows the same rule as the spectra of vowels.

Although it seems that adequate performance cannot be obtained since the proposed method is a rough transform, it was proven that the additional high band spectrum is very effective even if it is fixed type additional spectrum [5], [6]. It is expected in this case, since approximation is adaptive, to be more effective than conventional methods [5], [6].

3. IMPLEMENTATION

The block diagram of the proposed system is shown in Fig. 2. Band limited (telephone) speech with 8 kHz sampling, is filtered by LP analysis. The narrow band (8 kHz sampling) signal is analyzed using LP parameters. A spectrum envelope can be obtained from the LP parameters.

To obtain higher quality speech we utilize the residual error of the LP analysis. The residual LP analyzed signal is then expanded by an interpolator with 16 kHz sampling. The expanded residual signal is shaped by a shaping filter [4], [5]. Using LP coefficients, the expanded residual signal is applied to the LP synthesis filter. The high pass signal (> 3.4 kHz) and low pass signal (< 0.3 kHz) are added to the received telephone band (0.3-3.4 kHz) signal with 16 kHz sampling which reproduces the wide (upper) band signal.

In the signal summing part, all band signals are summed with suitable level adjustment in order to obtain a more suitable auditory evaluation. The level ratios are fixed to previously decided values.

In the extrapolation block, narrow band waveforms are analyzed by short term Fourier analysis, that is DFT (Discrete Fourier Transform), practically FFT is used. The slope of the spectrum envelope is obtained by smoothing the spectrum. This yields the slope of the spectrum envelope of the out-band signal.

Inverse DFT (IDFT) is carried out on the spectrum of the out-band signal to achieve the out-band signal in the time domain. Then LP coefficients correspond to the out-band signal can be computed.

Details of extrapolation processing are shown in Fig. 3. In Fig. 3 decimation is necessary to maintain the same number of points in IDFT as in DFT. In the expansion box, the out-band DFT spectrum is generated by linear extrapolation. The output signal of the IDFT is fed to the LP analyzer to obtain LP coefficients. To generate the wide band signal in the LP synthesis stage, an in-band signal is not always necessary because the in-band signal in this path is replaced by the received signal.

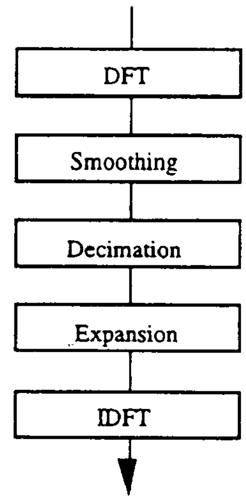

Fig. 3. Details of extrapolation processing

This proposed system does not use the conventional process of pitch extraction. This reduces the load for hardware implementation. It is also considered that the proposed system is more robust against noise.

It is known that for improving the speech quality in the high band accurate waveforms or spectra are not necessary, any signal with appropriate power can enhance the speech quality in the high band.

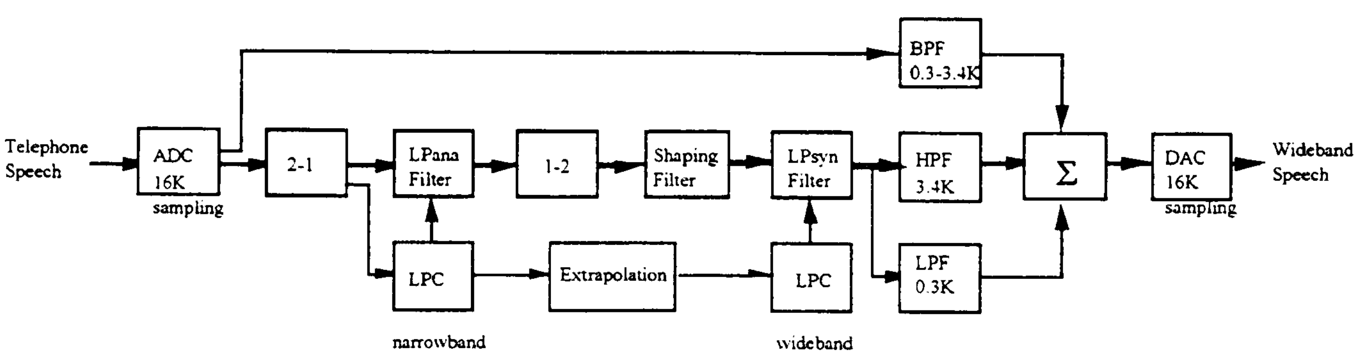

Fig. 2. Block diagram of the proposed system.

4. HARDWARE IMPLEMENTATION

Hardware implementation is also important. It has been shown that hardware based on DSP chips offers real-time performance. In practice, analog speech signals are converted in a 16 bit A-to -D converter with 16 kHz sampling frequency. Framing interval is 4 msec. Window length is 20 msec. Coefficient order in LP analysis is 15. Using those parameters the real time processing hardware can be implemented.

An example of the amplitude-frequency response of the shaping filter is shown in Fig. 4 The filter is a 13-th order FIR filter with a stop band frequency, 8 kHz, and a stop band attenuation, 60 dB.

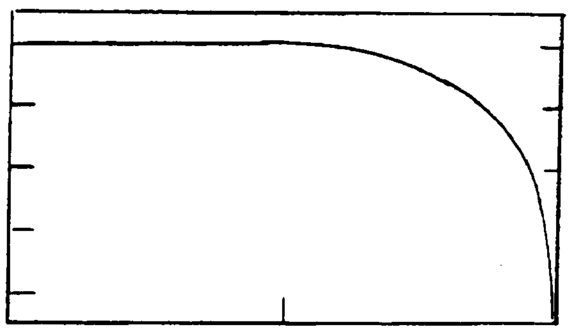

Fig. 4. Amplitude-frequency response of the shaping filter. (V: 20 dB/div)

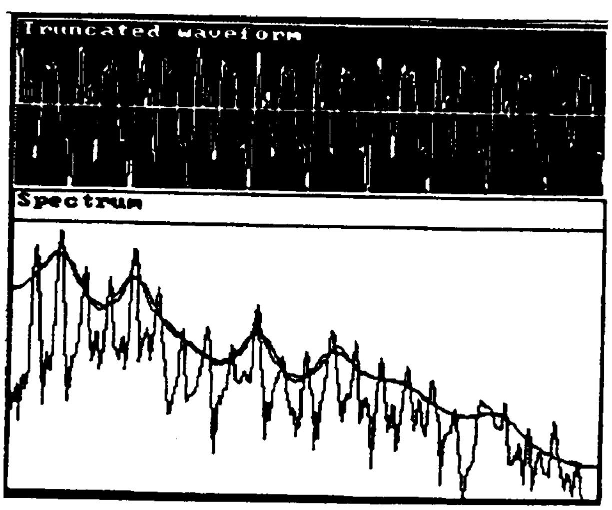

Fig. 5. Original speech (before transmitted) waveforms and spectrum.

5. PERFORMANCE

The waveforms and spectra of the original speech signal with 16 kHz sampling before transmission are shown in Fig. 5. The speech signal passes through the telephone network and is band limited in the telephone band as shown in Fig. 6.

The proposed system recovers the wide band signal by spectrum extrapolation. The spectrum of the speech signal output by the proposed scheme is shown in Fig. 7.

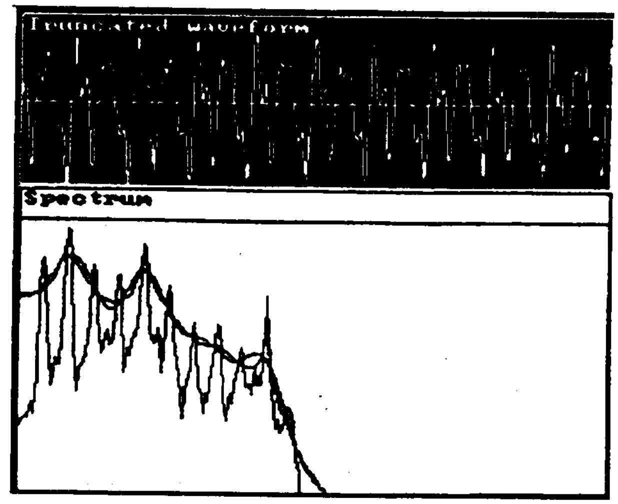

Fig. 6. Band limited (Telephone) speech waveforms and spectrum.

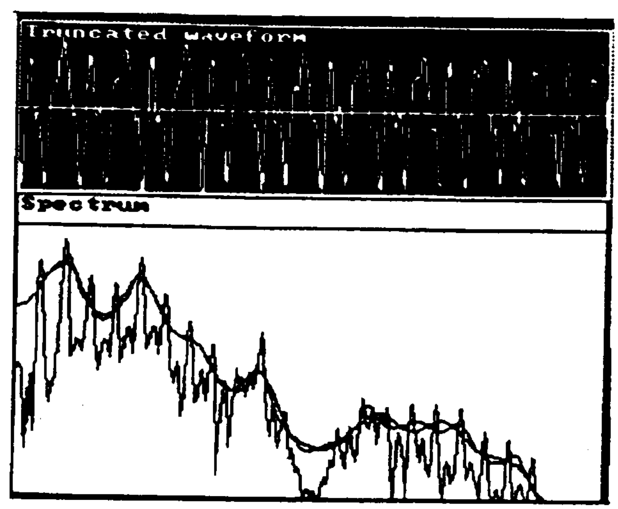

Fig. 7. Enhanced (recovered) wide band speech signal waveforms and spectrum.

The spectrum envelope of the output speech waveform is very similar to that of the original wideband speech signal. It is confirmed that the proposed scheme has superior performance in terms of lower output spectrum distortion and better speech quality compared to conventional methods.

By applying the shaping filter to the LP residual signal, the high band signal generated has better shape.

6. DISCUSSION

As the proposed method has a shaping filter for excitation input of the LP synthesis filter, the harsh sound characteristic of the LP synthesis filter can be suppressed. As spectrum extrapolation is made with linear smoothed slopes of the spectrum envelope, a more precise approximation can be achieved than by conventional fixed slope extrapolation. The proposed simple structure also helps to minimize the processing burden for hardware implementation.

As pitch extraction processing is not used in this proposed system, it offers good robustness against noise, complex speech or speech contaminated by noise etc.

In this scheme, vector quantization is not used, so distortion due to quantization error does not arise. Instead distortion due to approximation error must be considered. Of course, a final evaluation should be made by subjective assessment.

7. CONCLUSION

This paper has proposed a new implementation of the speech enhancement method that uses linear prediction analysis / synthesis and a simple but adaptive spectrum extrapolation method. Real time hardware implementation procedures were discussed, and its performance was clarified.

The method's performance evaluation against noise (robustness), and system evaluation by subjective assessments are future goals.

ACKNOWLEDGMENTS

The author wishes to thank Dr. T. Aoyama, Mr. I. Yamashita and Mr. K. Yoshida of NTT Optical Networks Systems Laboratories for their advice and encouragement.

REFERENCES

[1] CCITT recommendation G.722, Blue Book, Vol. III, Fascicle III.4, Geneva 1989.

[2] D. O. Bowker, et al., "Telephone network speech enhancement", US Patent No. 5195132, Dec. 1990.

[3] Y. M. Chen, et al., "Statistical recovery of wideband speech from narrowband speech", International Conference on Spoken Language Processing, ICSLP '92, pp. 1577-1580, 1992.

[4] Y. Yoshida, et al., "An algorithm to reconstruct wideband speech from narrowband speech based on codebook mapping", International Conference on Spoken Language Processing, ICSLP'94, 27.3, pp. 1591-1594, Sept. 1994.

[5] H. Yasukawa, "Quality Enhancement of Band Limited Speech by Filtering and Multirate Techniques", International Conference on Spoken Language Processing, ICSLP'94, 27.7, pp. 1607-1610, Sept. 1994.

[6] H. Yasukawa, "Enhancement of Telephone Speech Quality by Simple Spectrum Extrapolation Method", European Conference on Speech Communication and Technology, EUROSPEECH'95, pp. 1545-1548, Sept. 1995.

[7] H. Yasukawa, "Adaptive filtering for broad band signal reconstruction using spectrum extrapolation", The 7th IEEE 1996 Digital Signal Processing Workshop, 1996.

[8] C. Avendano, et al., "Beyond Nyquist: Towards the recovery of broad-bandwidth speech from narrow-bandwidth speech", European Conference on Speech Communication and Technology, EUROSPEECH'95, pp. 165-168, Sept. 1995.

[9] Y. Nakatoh, et al., "Recovery of wideband speech from narrowband speech based on piecewise linear mapping", (in Japanese) Technical report of IEICE, US95-76, EA95-70, Jan 1996.

[10] H. Yasukawa, "Signal Restoration of Broad Band Speech Using Nonlinear Processing", VIII European Signal Processing Conference, EUSIPCO'96, Sept. 1996.

[11] Y. Tanaka, et al., "Telephone speech enhancement by bandwidth expantion and spectral equalization", (in Japanese) Acoustical society of Japan Autum Convention report, 1-P-6, Oct 1994.

[12] H. Tasaki, et al., "Reconstruction of wideband speech from narrowband CELP codes", (in Japanese) Acoustical society of Japan Autum Convention report , 1-5-9, Oct 1994.

[13] K. Nagami, et al., "A method of generation of wideband speech from band-limited speech by autocorrelation function", (in Japanese) Acoustical society of Japan Spring Convention report , 2-P-10, March 1996.