Proceedings of The Institute of Acoustics

A RESIDUAL-EXCITED LPC ANALYSIS-SYNTHESIS SYSTEM FOR FORMANT-BANDWIDTH REDUCTION

K.K. Paliwal

Computer Systems and Communications Group Tata Institute of Fundamental Research Homi Bhabha Road, Bombay 400005, India

Abstract: Patients having hearing losses of cochlear origin are known to suffer from poor frequency resolution. It is conjectured that intelligibility of speech can be improved for such patients by reducing the formant bandwidths of the speech signal. In order to test the validity of this conjecture, we have developed a residual-excited LPC (linear predictive coding) analysis-synthesis system. An important advantage of this system is that it does not introduce any distortion of its own in the speech signal. Furthermore, it can be realized in hardware to operate in real-time with the present-day digital signal processors.

1. Introduction

Patients having hearing losses of cochlear origin are known to have broader auditory filters and, thus, suffer from poor frequency resolution [1,2]. . In fact, it has been shown physiologically for animals [3] and psychoacoustically for human patients [4] that frequency resolution becomes poorer with more severe cochlear hearing losses. It has been suggested that these hearing losses can be compensated for by sharpening of formant peaks in the short-term speech spectrum. That is, intelligibiling of speech can be improved for such patients by reducing the formant-bandwidths of the speech signal.

Summerfield et al. [5] used synthetic speech stimuli generated by a formant synthesizer to explore the effect of formant-bandwidth reduction. They found that formant-bandwidth narrowing did not improve the speech identification accuracy. They suggested two explanations for these negative results. The speech stimuli generated by the formant synthesizer were too unnatural and, hence, might have outweighed any advantage of improved formant resolution [5.6]. Alternatively, little may be gained by making formant-bandwidths narrower than the widths of patients' auditory filters. However, before passing the later verdict one should make the speech stimuli as natural as possible. This motivated us to undertake the present study.

Recently, Pick and Evans [1,2] have planned a similar study using an LPC analysis-synthesis system. Though the results of their study are not yet reported, they are also expected to have the same problem of unnatural speech stimuli, though to a less extent. The LPC analysis-synthesis system causes a degradation in speech quality (even when the formant-bandwiths remain unchanged) because of the following two reasons. Firstly, excitation to the LPC synthesis system is modelled as the pitch pulses for voiced speech and as random noise for unvoiced speech. This model is not perfect and, hence, causes some distortion in the resulting speech signal. Secondly, the LPC analysis-synthesis system requires voiced/unvoiced detection and pitch estimation which are prone to errors. These errors cause further distortion in the generation of speech stimuli in a conventional LPC analysis-synthesis system.

In the present paper, we try to solve these problems and propose a residual-excited LPC analysis-synthesis system for formant-bandwidth reduction. This system accepts real speech at its input, analyses it for LPC coefficients, transforms these coefficients for formant-bandwidth reduction, and uses the residual error signal as excitation for synthesising speech from the transformed LPC coefficients. Since the present system uses the residual signal itself as excitation, it does not introduce any distortion of its own in the resulting speech. Another advantage of the present system is that it is computationally more efficient because it does not require voiced/unvoiced detection and pitch estimation.

2. System description and some examples

Figure 1 shows the block diagram of the residual-excited LPC analysis-synthesis system. Here, a frame-wise LPC analysis of the speech signal is performed with a rate of 100 frames/s and a frame length of 20 ms. The Burg nethod is used for LPC analysis because it follows a lattice structure which is most suited for hardware implementation and always quarantees the stability of the LPC synthesis filter [8]. In addition, its performance is comparable to that of the autocorrelation and covariance methods for longer duration analysis frames (durations larger than two pitch periods) [9]. The LPC coefficients $\{a_k\}$ obtained from this analysis define the inverse filter

$$A(Z) = 1 + \sum_{k=1}^{p} a_k Z^{-k},$$

where p is the order of LPC analysis. The speech signal $\{s_n\}$ is processed by this inverse filter fo produce the residual error signal $\{e_n\}$.

In order to reduce the formant bandwidths, we use the following procedure: Solve the polynomial A(z) for its roots (by factoring it), compute formant-bandwidths, reduce them explicitly by the desired amount and compute the transformed LPC coefficients $\{a'_k\}$. In this procedure, factorization of A(z) is computationally expensive operation. However, Rajasekaran and Doddington [10] have recently shown that it is possible to factor in real-time the polynomial A(z) on Texas Instruments' digital signal processing chip TMS 320. The transformed LPC coefficients $\{a'_k\}$ define the LPC synthesis filter

A RESIDUAL-EXCITED LPC ANALYSIS-SYNTHESIS SYSTEM FOR FORMANT-BANDWIDTH REDUCTION.

$$H(Z) = 1/(1 + \sum_{k=1}^{p} a'_k z^{-k}).$$

This filter is excited by the residual signal (e_n) to generate speech (s'_n) which has reduced formant-bandwidths.

In order to see whether the residual-excited LPC analysis-synthesis system introduces any distortion of its own, we have processed speech utterances of two speakers (one male and one female) with a formant-bandwidth reduction factor of 1. As expected, the processed speech signal (s'_n) was found to be exactly identical to the original speech signal (s_n). Now we show two examples (one for male speaker and another for female speaker) of the speech signal processed by the present system with a formant-bandwidth reduction factor of 2. We show in Fig. 2 (a) a 32 ms segment of the male vowel sound /I/ (as in 'fifth') before (in solid line) and after (in dotted line) processing. As expected, the processed speech signal shows less damping between two pitch pulses than the original speech signal. We show in Fig. 2(b) the FFT spectra and in Fig. 2 (c) the LPC spectra. It can be seen here that the present system is effective in reducing the formant-bandwidths of the speech signal. We show in Fig. 3 a similar example for the female vowel sound /3/ (as in 'person').

We have used this system for reducing the formant bandwidths of a number of speech utterances of C-V syllables. A perception experiment is planned in collaboration with Dr. G.F. Pick of the Keele University, U.K., where the effect of formant-narrowing on the vowel identification accuracy will be studied for the patients having hearing losses of cochlear origin. The results of this experiment will be reported in our next paper.

3. Conclusion

A residual-excited LPC analysis-synthesis system for reducing the formant bandwidths is described. This system is computationally efficient and can be realized with present-day technology. Also, it is shown that this system does not introduce any distortion of its own. This system will be used to study the effect of formant-narrowing on the vowel identification performance for the hearing-impaired listeners.

References

[1] G.F.Pick and E.F.Evans, "Strategies for high-technology hearing aids to compensate for hearing impairment of cochlear origin", In "High Technology Aids for the Disabled", (W.J.Perkins, Ed.), Butterworth, London, 1982, pp.99-105.

[2] E.F.Evans, "Pathophysiology of the peripheral hearing mechanism". In "Hearing Science and Hearing Disorders",

(M.E. Lutman and M.P. Haggard, Eds.), Academic Press, London, 1983, pp. 61-80.

[3] E.F.Evans, "Normal and abnormal functioning of the cochlear nerve". In "Sound Reception in Mammals", (R.J.Bench, A.Pye and J.D. Pye, Eds.), Academic Press, London, 1975, pp. 133-165.

[4] G.F.Pick, E.F.Evans and J.P.Wilson, "Frequency resolution in patients with hearing loss of cochlear origin", In Psychophysics and Physiology of Hearing", (E.F.Evans and J.P.Wilson, Eds.), Academic Press, London, 1977, pp. 273-281.

[5] A.Q.Summerfield, R.Tyler, J.Foster, E.Wood and P.J.Bailey, "Failure of formant bandwidth narrowing to improve speech reception in sensorineural impairment", J. Acoust. Soc. Am., Vol.70, 1981, pp. S108-S109.

[6] P.J.Bailey, "Hearing for speech: the information transmitted in normal and impaired hearing", In "Hearing science and Hearing Disorders", (M.E.Lutman and M.P.Haggard, Eds.), Academic Press, London, 1983, pp. 1-34.

[7] M.P.Haggard; "New and old conceptions of hearing aids," In "Hearing Science and Hearing Disorders", (M.E.Lutman and M.P.Haggard, Eds.), Academic Press, London, 1983, pp. 231-282.

[8] J.Makhoul, "Stable and efficient lattice methods for linear prediction", IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-25, No.5, Oct. 1977, pp. 423-428.

[9] K.K.Paliwal and P.V.S.Rao, "On the performance of Burg's method of maximum entropy spectral analysis when applied to voiced speech", Signal Processing, Vol.4, No.1, Jan.1982, pp. 59-63.

[10] P.K.Rajasekaran and G.R.Doddington, "Real-time factoring of the linear prediction polynomial of speech signals", In "Digital Signal Processing-84", (V.Cappellini and A.G. Constantinides, Eds.), North-Holland, Amsterdam, 1984, pp. 405-410.

<u>Figure captions</u>

Fig. 1. Residual-excited LPC analysis-synthesis system for formant-bandwidth reduction

Fig. 2. Example of 32 ms male vowel sound /I/ processed for formant-bandwidth reduction factor of 2. Original speech

is shown here in solid line and processed speech in dotted line. (a) Waveforms, (b) FFT spectra and (c) LPC spectra.

Fig. 3. Example of 32 ms female vowel sound /3/ processed for formant-bandwidth reduction factor of 2. Original speech is shown here in solid line and processed speech in dotted line. (a) Waveforms, (b) FFT spectra and (c) LPC spectra.

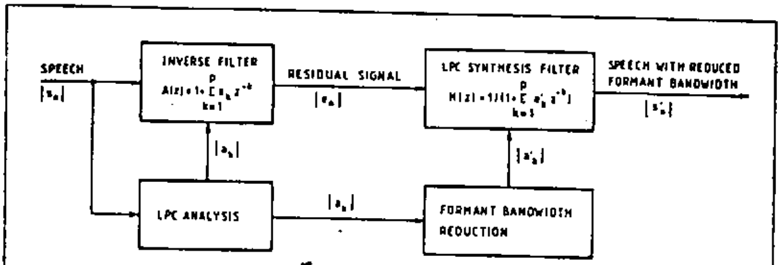

Fig. 1

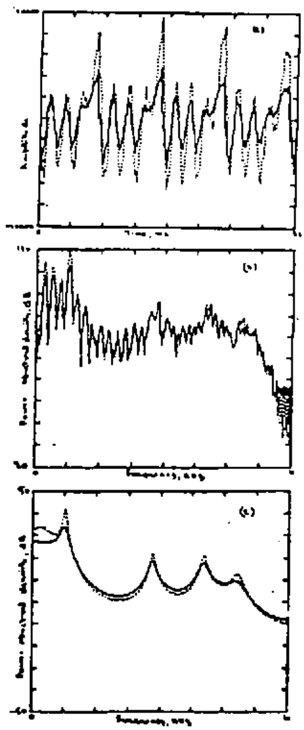

Fig. 2

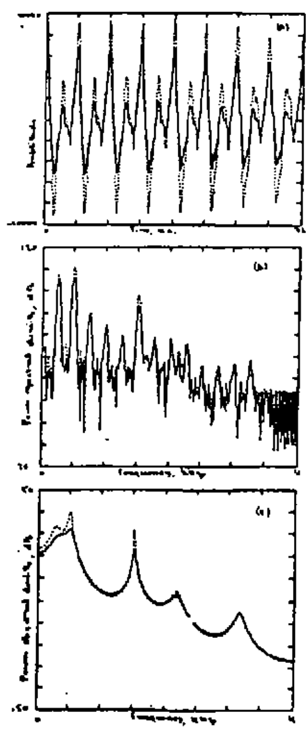

Fig. 3

COMPARATIVE EVALUATION OF LP ANALYSIS METHODS FOR MULTI-PULSE EXCITED LP CODING OF SPEECH

T. Svendsen and K.K. Paliwal

Division of Telecommunications
University of Trondheim, Trondheim-NTH, Norway

Abstract:

Multi-pulse excited LP (linear prediction) coding has been recently suggested as an alternative technique of producing natural sounding speech at medium bit-rates. However, in this coding technique the speech quality may depend on the type of LP analysis method used. In the present paper, we have studied standard methods of LP analysis, namely, the autocorrelation method, the covariance method, the modified autocorrelation method [2] and the Burg method. In terms of segmental signal-to-noise ratio, we have found the autocorrelation method and Burg's method to be the best.

Simulations:

The multi-pulse coder used in the simulations is based upon the original proposal by Atal and Remde [1], its block diagram shown in fig. 1. The mode of operation was the following:

The LPC-analysis was performed over blocks of 20 ms of speech, and the analysis parameters were updated every 10 ms. The overlap percentage was 50% in both forward and backward direction and the order of the LPC-analysis was 10. The block-length for the multi-pulse error minimization was 10 ms and the search for optimum pulse amplitudes and positions were performed over the entire 10 ms block. Each block was searched for 8 pulses, yielding a pulse/sample ratio of 1/10. The weighting filter was of the form

$$w(z) = \frac{1 - A(z)}{1 - A(\gamma z)}$$

where $1 - A(z)$ is the LPC prediction filter and $\gamma = 0.8$ is a weighting coefficient.

The input speech was two English sentences, one spoken by a man, the other spoken by a woman. Total length of the test sequence was 5.3 s. The sampling frequency was 8kHz.

Simulation results for natural speech are shown in Table I

LPC-algorithm	SNR [dB]	Segmental SNR [dB]
Mod. covariance	10.69	10.44
Covariance	5.72	10.51
Burg	10.75	10.86
Autocorrelation	10.63	10.98

Table 1. SNR-performance for the multi-pulse coder for different LPC-algorithms. Natural speech.

The cause for the low SNR value for the covariance method is instability in the LPC-filters.

Simulations were also carried out for telephone (IRS) filtered speech, and the results are shown in Table 2.

LPC-algorithm	SNR [dB]	Segmental SNR [dB]
Mod. covariance	9.65	7.34
Covariance	8.46	7.28
Burg	9.68	7.38
Autocorrelation	9.79	7.64

Table 2. SNR-performance for the multi-pulse coder for different LPC-algorithms. Telephone speech.

The drop in SNR-values are due to the multi-pulse algorithm's ability to better trace low-frequency components containing high energy than high-frequency components.

We also simulated the Burg algorithm when performing a joint optimizatien of the pulse amplitudes after each new pulse was found [3]. For natural speech the simulation resulted in SNR = 11.28 dB, segmented SNR = 11.08 dB. The corresponding figures for telephone speech were 10.07 dB and 7.57 dB respectively.