Sindh Univ. Res. Jour. (Sci.Ser.), Vol: 34-B (2) 29-34 (2002)

PACKETIZED SPEECH QUALITY IMPROVEMENT WITH BANDWIDTH EXPANSION APPROACH

Asadullah Shah and Salma Bhatti

Department of Computer Science, Isra University, Hyderabad

Abstract

Speech packets can be corrupted while travelling through a transmission medium. The corruption of packet causes quality impairments. Speech quality, coded at low bit rates, is rated below excellent quality. Therefore, additional loss of packets further causes degradation in speech quality. There are lost frame reconstruction techniques through which the effects of lost and corrupted packets, on speech quality can be mitigated. Bandwidth expansion is one of the approaches applied upon Linear Prediction Coding (LPC) based packetization to improve speech quality. In this paper, a Codebook Excitation Linear Prediction (CELP) system of 4.3 kb/s speech is packetized and a random packet loss phenomenon is simulated. The lost packets are reconstructed by pitch filter and LPC filter with bandwidth expansion of LPC coefficients. The simulation results for various loss rates and corresponding speech quality measures are presented with Mean Opinion Scores (MOS). It has been observed that with bandwidth expansion, quality is improved over without bandwidth expansion schemes in similar scenario.

Keywords: Packetized speech quality, Bandwidth expansion approach.

Introduction

Digital speech coding techniques have been existing since early 1960's. Since then speech coding and compression field has been getting lot of attention especially for the reduction of bit rates and maintaining the speech quality. After PCM [1], Adaptive Differential PCM (ADPCM) [2] was introduced at half the bit rate, i.e. 32 kb/s. With the passage of time, speech compression techniques are developed that are operating at rates as low as around 2kb/s [3][4]. The research is being carried out to develop techniques for further rate reduction, such as, Codebook Excited Linear Prediction, its variants and prototype speech coding techniques.

Experiments for speech packetization at higher bit rates have been conducted through data communication networks. But, at low bit rates, the research is now picking up, [5][6], yet the major obstacle for low bit rate packetization is the quality of speech. In this paper the issue of lost packets, reconstruction of those packets with subjective measures is presented.

Packet Formation

Low bit rate coding techniques are based on sequential coding of segments commonly known as frames of speech. The choice of the size of these segments determine the bit rate and the quality of speech. For Code-book Excitation Linear Prediction (CELP) type speech algorithm [7][8], three sets of parameters are used to encode and decode each segment of speech. These parameters are then digitally coded directly or transformed into frequency domain and then coded and bit stream of digital codes, frame by frame (segment) are transmitted. At the receiving side of the system, a similar algorithm is activated to reverse the process done at the transmitter and each frame is recovered back to produce original speech signal. As part of the algorithm requirement, the coding parameters of each frame are restored at both ends for encoding and decoding the next frame.

CamScanner

For any lost packet the parameters of a previous frame are exploited to mitigate the effects of those packets on speech quality.

Simulation Parameters

For simulation speech encoder [8] at 4.3 Kb/s is used, three packet sizes are considered (a) packet of single frame of speech, in this case is 30 ms (b) packet of two frames of speech (60 ms) and (c) a packet of three frames of speech (90 ms). The loss of packets is simulated on a random basis. All the packets are equally probable for the loss and no priority is given to any packet, such as, talkspurts and silences are also considered, and no Voice Activity Detection (VAD) is included for simulations.

The lost packets are reconstructed at the receiver, with bandwidth expansion approach. The bandwidth expansion can only be applied upon Line Spectral Frequencies (LSF's) commonly known as Short Term Prediction (STP) or LPC coefficients. There are LPC-10 (ten such coefficients) in proposed CELP of 4.3 Kbps system. The long-term prediction LTP parameters are reconstructed as given

$$W(z) = \frac{A(z)}{A(z/\gamma)} = \frac{1 - \sum_{j=1}^{P} \alpha_j z^{-j}}{1 - \sum_{j=1}^{P} \alpha_j \gamma^j z^{-j}}$$

---------- Equ. 1

Where α_j are coefficients of the LPC filter of order p and γ is the factor that controls the error energy in the formant regions as shown in Fig. 1. It is commonly referred to as the noise spectral shaping or, in short, noise shaping factor and its value lies between 0 and 1. Decreasing the value of γ broadens the formats. The increase in formant bandwidth is given by

[9]. The third set of parameters; CodeBook (CB) [10] are ignored because of their least contribution and effect on speech quality.

Noise shaping filter and bandwidth expansion

The most commonly used error criterion in AbS-LPC [11] techniques for computing the LPC parameters and optimum excitation is the minimum mean squared error (MMSE) mainly because of its simplicity. It offers adequate performance at high (> 16kb/s) bit rates. However, at lower rates (<8 kb/s) where the availability of merely one or less bits per sample makes it more difficult to reconstruct a waveform, which closely matches the input waveform, the MMSE is not much meaningful. Hence, an error criterion based on speech perception rather than waveform matching is required. Unfortunately, a single perceptual error criterion, which offers a totally satisfactory performance, has not yet emerged. In the mean while therefore, MMSE with a weighting filter [12],[13] has been commonly used AbS-LPC technique. The transfer function of this weighting filter is given as,

$$\Delta f = -\frac{f_s}{\pi} \ln \gamma \quad Hz$$

---------- E.q. 2

Where f_s is the sampling frequency. The optimum value of γ is selected empirically by subjective listening tests. For speech sampled at 8 kHz, it usually lies in the range of 8.0 to 0.9.

CamScanner

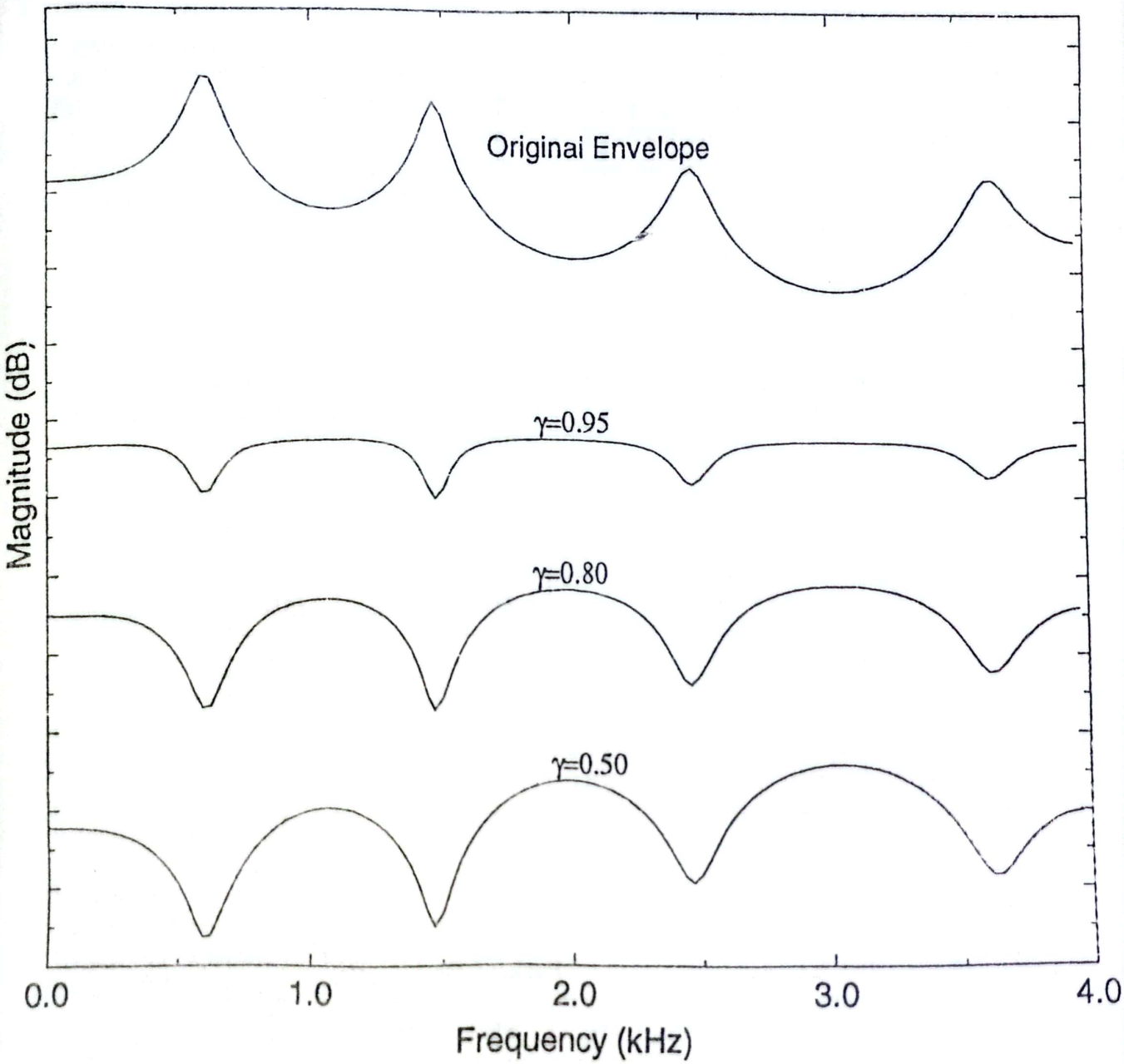


Fig. 1: LPC Envelop and Noise Shaper Filter

CamScanner

The basic philosophy behind the application of the weighting filter is based on the well known "masking" property of the ear, i.e. low energy signals are masked by high energy signals in a given frequency band. In other words, the ear is less sensitive to noise in the formant regions as compared to the valleys. Thus, by shaping the otherwise approximately flat noise spectrum during the search for the optimum excitation in such a way that the formants receive less attention (allowing for higher error energy), overall perceptual distortion can be minimized which results in higher subjective quality.

Mean opinion score (mos) results

Normally LSFR schemes apply post-pitch filter to emphasize pitch filter parameters. The LSFR [9][10] reported exploit the pitch parameters and substitute previous frame LTP coefficients to reconstruction lost frame. The bandwidth expansion, as reported in previous sections is only applied upon noise shaping filter to de-emphasize the regions of signal where noise is less effective for human perception. Comparisons of MOS results of without, and with bandwidth expansion are given in Table 1 (a) and (b) respectively.

Size	Packet Loss				
	0%	1%	5%	10%	15%
30 ms	3.5	3.1	2.73	2.18	1.9
60 ms	3.5	2.7	1.9	1.6	1.4
90 ms	3.5	2.3	1.5	1.4	1.2

(a)

Size	Packet Loss				
	0%	1%	5%	10%	15%
30 ms	3.5	3.2	2.8	2.3	1.9
60 ms	3.5	2.8	2	1.7	1.6
90 ms	3.5	2.4	1.6	1.5	1.4

(b)

Table 1: MOS (a) without (b) with bandwidth expansion

CamScanner

It is shown by the MOS test results that bandwidth expansions for all the packet loss rates for all the packet sizes perform slightly better then without bandwidth expansion, as shown in Table 1.

Discussion

The loss of speech packets of 3 sizes are considered to be dropped at random basis. Irrespective of speech signal a packet is dropped and reconstructed at the receiver using LSFR scheme. As shown in Table-1 (a) and (b), when there is no loss of packets, speech quality is rated 3.5 that means good quality. For 1% packet loss, the quality is rated better for bandwidth expansion schemes. Similarly for 5%, 10% and 15% losses speech quality is rated higher as perceived by the subjects (Listeners) than that of bandwidth expansion free scheme.

Conclusion

The smaller size packets are better to be reconstructed due to the fact that the previous frame parameters can substitute better for next single frame. But the longer size packets of size 2 and 3 frames the speech signal has traveled a distance of 60-90 ms apart. The reconstruction parameters of a lost frame for these packets can not provide better substitute. Bandwidth expansion in addition to pitch and post-pitch filtering perform better then pitch and post-pitch filter without bandwidth expansion.

References

[1] Rabiner L. and Schafer R., "Digital Processing of Speech Signal". Signal Processing, Prentice-Hall, 1978.

[2] Jayant N. and Noll P. "Digital Coding of Waveforms: Principles and Applications to Speech and Video", Signal Processing, Prentice-Hall, 1984.

[3] Kondoz, A.M. "Digital Speech Coding for Low Bit Rate Communicatiopn Systems", John Wiley, Chichester, UK, 1994

[4] Alkinson A., Kondoz A.M. and Evans B.G., "Time Envelope Vocoder, A New LP Based Coding Strategy For Use at Bit Rates of 2.4 Kb/s and Below", IEEE Journal on Selected Area in Communication, Vol. 13, No. 2, 1995, 449-457.

[5] Clifford J. Weinstein and Janes W. Forgie, "Experience with Speech Communication in Packet Networks", IEEE Journal on Selected Area in Communication, Vol. SAC-1, No. 6, Dec: 1983, 963-980.

[6] Shah, A. Ansari A. W. and Lubna Mustafa, "Low bit rate Speech Packetization and Packet Reconstruction", Mehran Univ: Res· Jou: of Eng: and Tech:, Vol 1ᶜ, No.1, Jan: 2000, 9-12.

[7] Mustafa, L. Shah, A. Khoumbate K.R. and.Kondoz, A.M " Packetized Speech Quality Improvement by Post-Pitch Filtering Approch", Sindh Univ: Res:Jou:(Science Series), Vol. 31, No.1, 1999, 25-38.

[8] Schroeder M. and Atal B.S, "Codebook Excited Linear Prediction (CELP): High Quality Speech at Low Bit Rates", In Proc. of ICASSP, Florida 1995, 937-940.

[9] Kondoz, A.M., Horos J. Evan B.G., and Suddle M.R., "Pulse Residual Excited Linear Prediction", IEE Proc. on Vision and Image Processing, 192 (2), April 1995, 105-110.

CamScanner

[10] S.A. Atugsiri, R.Soheili, A.M.Kondoz and B.G. Evans., "EffectiveLost Speech Frame Reconstruction for CEELP Coders", In Proc. of EuroSpeech Conf, Volume 2, Geneva, Italy, Sep: 1991, 599-602.

[11] Salma Bhatti, Shah, A. "Parameter based Low Bit Rate Speech Packetization" accepted, Sindh Univ: Res: Jou: (Science Series), Vol. 34, No. 1, 2002.

[12] ONG L.K. "Source Reliant error Control For Low Bit Rate Speech Communications", Ph.D. Thesis, University of Surrey, Guilford, U.K., 1994.

[13] Suddle M.R. "Speech Coding is Private and Broadcast Networks", Ph.D. thesis, University of Surrey, Guilford, U.K., 1996.

CamScanner