A NEW APPROACH FOR ENHANCING SPEECH SIGNALS

M.Thillaikarasi
Assistant professor
Department of computer science and Engineering
Annamalai University
Annamalai Nagar

INTRODUCTION

This paper proposes a novel approach that a system which enhances the speech signals, encodes it using bit rate. The noisy speech signals are being enhanced through speech enhancer , which is obtained by using code excited linear prediction [CELP]. This provides good scalability and reliability. Further the encoded speech gets compress voice signal using a lossy compression technique called linear predictive coding (LPC) vocoder, this is again the system of encoding speech at a low bit rate. To create an excitement signals in the decoder, the encoder sends voiced\unvoiced information, and the pitch period for voiced segment which is used. Now, the encoded gets decoded and the signal is reconstructed. The performance of the system is attained by the reconstructed speech. This inference is carried out and performance shows increase in lucidity of the reconstructed signal. The reconstructed signal evidences the efficiency of the project.

Keywords Code Excited linear predictor,linear predictor vocoder

1.1 Speech Enhancement

Speech enhancement aims to improve speech quality by using various enhancement algorithms. The objective of enhancement is improvement in intelligibility and/or overall perceptual quality of degraded speech signal . The applications of speech enhancement include mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids. A practical speech enhancement system generally consists of two major components: the estimation of noise power spectrum, and the estimation of speech [1].

1.2 Speech Coding

Speech coding or compression is a process of obtaining a compact representation for the speech signals, for the purpose of efficient transmission over band limited wired or wireless channels and also for efficient storage. In recent day's speech coders became the essential components for telecommunications and multimedia as the utilization of the bandwidth affects the cost of transmission. The goal of speech coding is to represent the samples of a speech signal with a minimum number of bits without any reduction in the perceptual quality. Speech coding helps a telephone company to carry out more voice calls on a single fiber link or cable. Speech coding is also useful for Voice over IP, Video conferencing and Multimedia applications to reduce the bandwidth requirement over internet. In addition, most of the speech applications require minimum coding delay because long coding delays hinder the flow of the speech conversation. Speech coding has been and still a major issue in the area of digital speech processing.

1.2.1 CELP system block diagram

The CELP Analysis-by-Synthesis system is as shown in Fig.1 where encoding and decoding of speech takes place at the encoder and the parameters which minimize the energy of error signal are found at the encoder. LP analysis is used to find the vocal system impulse response in each frame. The error signal is perceptually weighted to emphasize important frequencies and it is minimized by optimizing the excitation signal. The excitation signal is updated over four blocks within the frame. The proposed CELP coder has a frame duration of 20ms and block duration of 5 ms for finding the excitation. The encoder needs information about linear prediction coefficients (a), gain (G), pitch filter (b), pitch delay (P) and codebook index (k). After calculating these parameters they are sent to decoder. The linear prediction analysis estimates all pole filter in each frame which is used to generate spectral envelope of the speech signal. The filter has 10 LP-coefficients and makes use of Levinson‟s Durbin algorithm which reduces the complexity of the filter. The output of LP Analysis is error signal which is passed through the perceptual error weighing filter to control the noise level. The Gaussian codebook in the implemented system has number of Gaussian signals which are used as excitation signals for the filter. The Gaussian codebook with 512 sequences yields good quality speech with code index value as 9 bits. The pitch filter is used as a long delay correlation filter to generate pitch periodicity in the voiced speech. For energy minimization between original speech signal and synthetic speech the parameters G, k, b and P are determined over a particular frame [12].

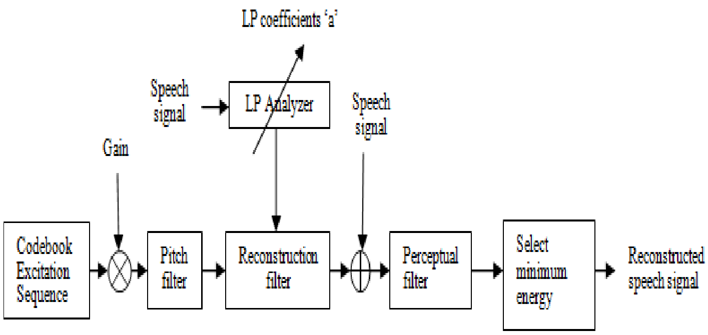

Fig 1.block diagram of CELP

LP analyser

The linear prediction analysis estimates the all-pole (vocal-tract) filter in each frame, used to generate the spectral envelope of the speech signal. The filter typically has 10-12 coefficients. In our implementation it has 10 coefficients. We have used MATLAB's lpc function to obtain these coefficients however they can be obtained by implementing a lattice filter which acts both as a forward and backward error prediction filter. It gives us reflection coefficients which can be converted to filter coefficients. Levinson-Durbin method can be used effectively to reduce complexity of the filter.

Pitch filter

Human voices have pitch in a few hundred hertz .In our implementation we consider pitchfrequencies from 50Hz to 500Hz. For 8 kHz signal these frequencies correspond to pitch delay of 16 to 160 samples. For voiced speech, the excitation sequence shows a significant correlation from one pitch period to the next. Therefore, a long-delay correlation filter is used to generate the pitch periodicity in voiced speech.

Excitation sequence

The codebook contains a number of Gaussian signals which are used as the excitation signals for the signal e(n) used to excite the LP synthesis filter 1/A(z) is determined every 5 milliseconds literature we find different types of codebooks. Most widely used are adaptive codebooks in itch Filter uman voices have pitch in a few hundred hertz. In our implementation we consider pitch T the filer. In our implementation we generated a codebook of 512 sequences each of length 5ms i.e. 40 samples. The codebook is known to the encoder as well as the decoder. T within the frame under analysis. An excitation sequence dk(n) is selected from a Gaussian codebook of stored sequenced, where k is the index. If the sampling frequency is 8 kHz and the excitation selection is performed every 5ms, then the codebook word size is 40 samples. A codebook of 512 sequences has been found to be sufficiently large to yield good-quality speech, and requires 9 bits to send the index. In conjunction with the fixed codebook. However we stick to fixed codebook which is a collection of Gaussian signals

Energy minimization

The excitation sequence e(n) is modeled as a sum of a Gaussian codebook sequence dk(n) and a sequence from an interval of past excitation.

1.2.2 Post filtering in CELP

The perceptual weighting filter is used inside search lop for best excitation in the codebook. When there is some distortion remaining in the reconstructed speech, it is termed as roughness or coding error which is a function of frequency and too high at regions between formants and between pitch harmonics. Thus several coders employ a post filter that operates on reconstructed speech to de-emphasize coding error between formants and pitch harmonics. This process is known as “Post-processing”

.

Application of Digital hearing aids

Digital hearing aid systems consist of different configurable signal processing blocks. Significant research has been devoted to the signal processing domain for hearing aids and many different algorithms were suggested. [6] have listed the recent achievements of different blocks and suggest that the critical signal processing blocks of today’s high-end digital hearing aid include adaptive feedback cancellation, adaptive beamfomer, filterbank, adaptive noise reduction and wide dynamic range compression. Based on different application requirements and resource budget, the blocks and algorithms for these blocks vary in different hearing aid systems as shown in the figure..

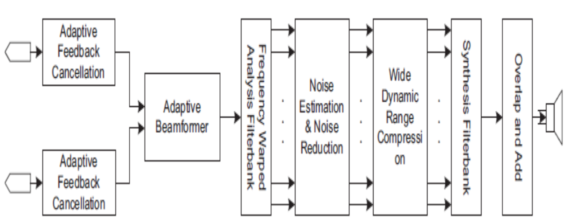

Fig 2 block diagram of digital hearing aids system

Feedback is caused by the re-amplification of output from the speaker. A feedback path model of the environment must be established in order to cancel the feedback. Generalized sidelobecanceller[7] is a typical adaptive beamformer algorithm in hearing aids. It suppresses signals from the sides and rear using an array of two omni-directional microphones.and Frequency warped filterbank (FWFB) This block uses a warped FIR filter to approximate the frequency resolution of the ear[7]. It is obtained by replacing the unit delay in a digital filter with first-order all-pass filters . In this design, the signal is divided into 17 frequency bands.

Noise Reductionfilter*(NRF):*The major anxiety for the people with hearing loss is the capability of hearing aids to differentiate intended speech signal in a noisy environment. Hence, to eliminate the noise, a reduction filter function is used in this design. The ears of hearing-impaired people have a smaller dynamicrange than those of healthy people. This block applies different gains for different input levels to map thenormal dynamic range to the reduced range

OBJECTIVE PERFORMANCE MEASUREMENT

In this section, the Mean Opinion Score (MOS) is used to evaluate resynthesized sounds [8]. A group of five people is asked to rate the quality of synthesized sounds based on a scale of 1 to 5 (1 = Bad, 2 Poor, 3 = Fair, 4 = Good, 5 = Excellent). Fig.7 below shows the MOS values of LPC for female, male, and bird sound recordings. The scores are evaluated by five volunteers who allowed using their voice recordings in this project.

Test condition	**MOS**
Noisy speech	3.09
Noisy decompressed speech	3.05
Enhanced decompressed speech	3.08
Male original speech	4.15
Male decompressed speech	3.92
Male enhanced decompressed speech	4.10
Female original	3.98

speech	
Female decompressed speech	3.90
Female enhanced decompressed speech	3.96

Table 1.MOS

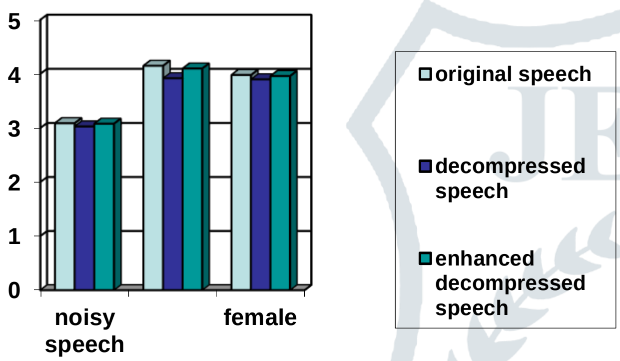

Fig 3 MOS

EXPERIMENTS AND RESULTS

This report discusses a study of a new type of speech coding algorithm that in many ways fills a void in the capabilities of present generation speech coders. Previously studied waveform coding schemes tend to have a knee in the speech quality / bit rate curve such that for rates substantially below 10 kb/s, the quality of the reproduced speech falls off rapidly. At rates below 3 kb/s, vocoders which produce synthetic quality speech are the only alternative. The Code Excited Linear Prediction (CELP) coding scheme studied here tends to fill in the gap between waveform coders and vocoders at rates around 5 kb/s. Rates around 5 kb/s are of practical importance. While bandwidth is less of an issue in new digital services based on fiber optic transmission, a large class of applications still is heavily bandwidth limited. Two examples of applications for 5kb/s coders are for secure voice transmission over existing analog facilities, and in wide-area voice communications. The secure voice terminal application combines a low bit rate speech coder, a digital encryption device, and a data modem. Practical considerations limit the reliable full-duplex capabilities of the switched telephone net-work to rates around 5 kpbs. This provides the impetus for developing speech coders that produce good quality speech at this data rate. The second application involves the use of radio systems with relatively wide coverage. Examples are cellular systems for both mobile and fixed applications and wide coverage satellite systems for mobile and fixed voice applications. In both cases, bandwidth is a scarse resource. Due to the use of a wide coverage medium, digital encryption is also desirable. For these applications, a range of bit rates is possible. However, at rates near 5 kb/s, digital systems again become competitive in bandwidth with analog systems, while maintaining the benefits of a digital implementation. The complexity of the schemes being considered is at the high end of the scale for speech coders. In spite of this, the hardware complexity for the schemes studies can be measured in terms of one or at most a small number of the next generation of the DSP chips. The work described in this report has been aimed at producing good quality speech at bit rates around 4800 b/s. At these rates efficient coding of the side information becomes important. Previous studies of CELP have concentrated on coding the waveform information and synthesizing the speech with encoded side information. The waveform information rate is in fact the smaller part of the data flow that has to be sustained by our coder. The work on CELP has been supported by both the Communications Research Center and the Natural Sciences and Engineering Research Council (NSERC).

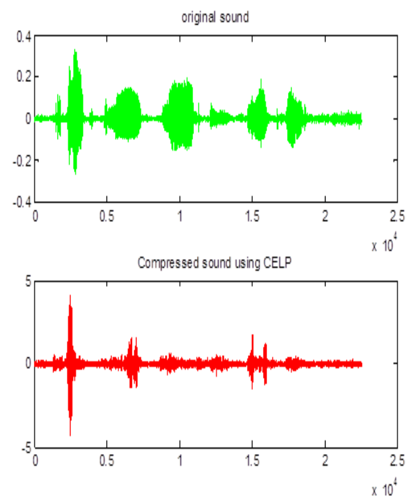

Fig 4.1 Sample Speech Signals

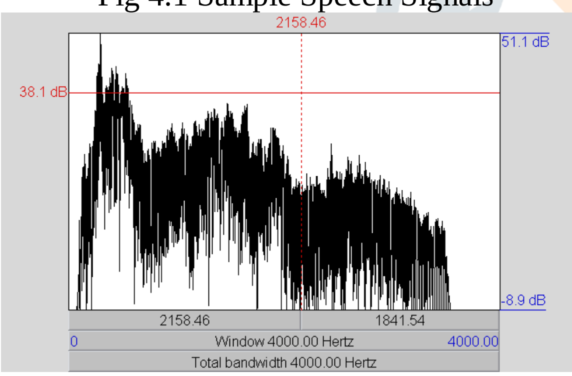

Fig 4.2 spectrum of speech signal

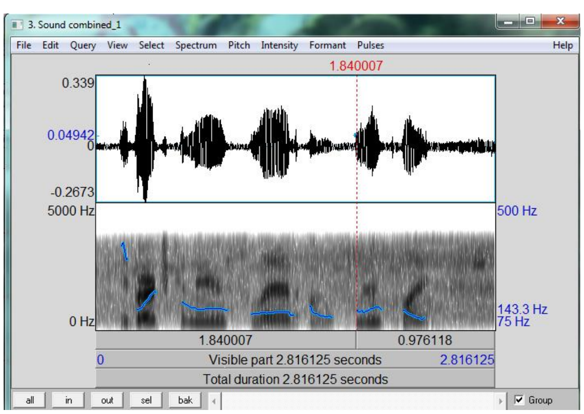

Fig 4.3 Spectrogram Of Speech Signal

Figure 4.1 shows the output of speech signal and having the original and compressed signals. Figure 4.2 shows the spectrogram of same signals.

CONCLUSION

So far, this blog has forced on designing a simple CELP coder which induces on good quality speech at low bit rate. Its performance on the various rates is noted. Enhanced speech is enhanced by using of CELP. To minimize its complexity “simple codebooks” were used and their performance was evolved.

LPC encoders fragment a sound signal into different segments and then send the information on each segment to the decoder. The encoder transmits the information, whether the segment is voiced or unvoiced and the pitch period for voiced segment. The developed excitement signal in the decoder encodes and also transmits information about the vocal tract; this would build a filter on the decoder side. This will only reproduce the original speech when the excitement signal is given as input. Enhancement techniques will be merged along with coder and quality of digital hearing aids can be improved.

References

1. Speech enhancement for non-stationary noise environments Israel Cohen ∗, Baruch Berdugo Lamar Signal Processing Ltd., P.O.Box 573, YokneamIlit 20692, Israel Received 18 February2001; received in revised form 26 June 2001
2. M. Hasegawa-Johnson, and A. Alwan, “Speech coding: Fundamentals and applications,” w*iley Encyclopedia of Telecommunications*, vol. 5, pp.
   2340-2359, 2003.
3. Hui Dong, Jerry D. Gibson universal successive refinement of celp speech coders, Electrical Engineering Department Southern Methodist University

4. European Telecommunications Standards Institute “Coding of speech at 16 kbit/s using low-delay code excited linear prediction,”, ITU –
   Recommendation G.728, Sep. 1992.
5. J. Campbell, Jr., T. E. Tremain, and V. C. Welch., “The federal standard 1016 4800 bps CELP voice coder,” Digital Signal Processing, vol. 1, no. 3, pp. 145–155, 1991
6. (2011). [Online]. Available: fhttp://speex.orgg
7. ITU-T P.800 Recommendation, “Methods for subjective determination of transmission quality,” 1996
8. Guilin Ma, *Student Member, IEEE*, Fredrik Gran, Finn Jacobsen,and Finn Thomas Agerkvist, “Adaptive feedback cancellation with band-limited lpc vocoder in digital hearing aids” IEEE Transacations On Audio, Speech, And Language Processing, MAY 2011
9. *Peter Kabal “Ill-Conditioning and Bandwidth Expansion in Linear Prediction of Speech” Proc IEEE Int. Conf. Acoustics, Speech, Signal Processing* (Hong Kong), April 2003
10. Arpana Mishra , Javed Ashraf “speech compression with voice excited linear predictive coding” *International Journal of Emerging Technology and Advanced Engineering, Website: www.ijetae.com , june2012*
11. Omesh Wadhwani*, Amit Kolhe, Sanjay Dekate “Recognition of Vernacular Language Speech for Discrete Words using Linear Predictive Coding Technique” International Journal of Soft Computing and Engineering (IJSCE) November 2011

12. Scott C. Douglas. Introduction to Adaptive Filters, Digital Signal Processing Handbook 1999
13. othman O.Khalifa, M.H.Makhtar, and M.S.Baharom " hearing aids system for impaired peoples" international journal of computing &information science , april 2004
14. peng qiao and henk corporaal, menno lindwer, 'A 0.964mW digital hearing aid system" 2011
15. Fa-Long Luo, Senior Member, IEEE, Jun Yang, Senior Member, IEEE, Chaslav Pavlovic, and Arye Nehorai, Fellow, IEEE, "Adaptive Null-Forming Scheme in Digital Hearing Aids, 2002