International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Implementation of Speech Enhancement Techniques Using DSP Processor

Kakularam Shobha Rani [a], Yadama Sankeerth [b], P S Sreenivas Reddy [c]

[a] Student Nalla Narasimha Reddy Education Society's Group of Institutions (Autonomous), Affiliated By Jntuh Hyderabad, Department Of Electronics and Communication Engineering , Hyderabad, Telangana, Pin Code 500088, India

[b] Student Nalla Narasimha Reddy Education Society's Group of Institutions (Autonomous), Affiliated By Jntuh Hyderabad, Department Of Electronics and Communication Engineering , Hyderabad, Telangana, Pin Code 500088, India

[c] Under the Guide, ASSOCIATE PROFESSOR, Nalla Narasimha Reddy Education Society's Group of Institutions

A B S T R A C T

In recent years, communication infrastructures have changed dramatically to include voice, data, image, and video services. However, speech is still the most common and fundamental service provided by the telecommunication networks. With an increased bandwidth to preserve audio signals fidelity, wideband coding is also gaining popularity with improved voice quality. This paper introduces speech coding techniques, speech enhancement methods, and voice over internet protocol (VOIP) applications for next generations networks. Excessive noise levels will also degrade the performance of the existing signal processing techniques, such as speech coding, speech recognition, speaker identification, and adaptive echo cancellation, which are developed under the low-noise assumption. Noise reduction becomes an increasingly important requirement for improving voice quality in noisy environments for hands-free applications.

Keywords: Digital signal processor, Speech Enhancement, Real time.

1. Introduction

In computer speech recognition, a person speaks into a microphone or telephone and the computer listens. Speech processing is the study of speech signals and the processing methods of these signals. The signals are usually processed in a digital representation. So speech processing can be regarded as a special case of digital signals processing applied to speech signals. Automatic Speech Recognition technology has advanced rapidly in the past decades. The efficiency of these methods depends on the intelligibility and the quality of the enhanced speech signal. The augmentation in the SNR (Speech signal-to-Noise Ratio) is the aim of most methods. Spectral Subtraction (SS) is the first approach for enhancing speech corrupted by additive noise. It approxi-mates the spectrum of the original speech signal via the subtraction of the magnitude spectrum of the estimated noise from the noisy speech signal magnitude spectrum, whereas keeping the phase spectrum of the noisy speech signal.

The shortcoming of this approach is the residual noise. Digital signal processing (DSP)is concerned with the digital representation of signals and the use of digital systems to analyze, modify, store transmit, or extract information from these signals. In recent years, the rapid advancement in digital technologies has enabled the implementation of sophisticated DSP algorithms for real-time applications. DSP is now used not only in areas where analog methods were used previously, but also in areas where analog techniques are very difficult or impossible to apply. Speech coding is the digital representation of speech signals in to digital codes and decompress the digital codes to reconstruct the speech signals to a satisfactory quality. Several sophisticated speech coding algorithms have been developed to preserve speech quality while achieving low bit rate. These algorithms usually require higher computational load and more memory for complicated programs and larger signal buffers. The trade-offs among the bit rate, speech quality, coding delay, and algorithm complexity are the main considerations when designing speech systems. The bit-rate reduction to 32kbps can be achieved by using the adaptive differential PCM (ADPCM) algorithm, which uses an adaptive predictor and differential quantizer to track the changing amplitudes of the speech signals.

2. SPEECH CODING TECHNIQUES

Speech coding is the digital representation of speech signals to provide efficient transmission and storage of digital data. The simplest method to encode speech is to uniformly sample and quantize (digitize) the time-domain speech waveform for digital representation, known as pulse code modulation (PCM). Most telecommunication systems use the 8KHz sampling rate, PCM coding requires a bit rate of 96 kilobits per second. Further bit rate reduction to 32kbps can be achieved by using the adaptive differential PCM (ADPCM) algorithm, which uses an adaptive predictor and differential quantizer to track the changing amplitudes of the speech signals. The basic concept of differential quantization is to quantize the difference between the speech sample and its prediction. Linear PCM, nonlinear companding (µ-law or A-law), and ADPCM are classified as waveform coding techniques, which operate on the amplitude of speech signals on a sample-by-sample basis. In contrast, the analysis-by-synthesis coding methods process signals on a frame-by-frame basis to achieve a higher compression rate by analyzing and coding the spectral parameters that represent the speech production model. The analysis-by-synthesis coding method transmit the coded parameters to the receiver for speech synthesis. This type of coding algorithm is called a vocoder (voice coder) since it uses an explicit speech production model. Many vocoders are based on linear predictive coding (LPC) techniques to achieve low bit rates. The LPC coding techniques will be introduced in this paper.

2.1 Speech production model using LPC

The LPC method is based on the speech production model including the excitation input, gain, and vocal-tract filter. The simplified speech production model for an LPC vocoder is illustrated in Figure 1. As introduced in the IIR filter design, amplification to certain frequencies can be achieved by properly placing the poles in the filter transfer function. The LPC synthesis filter is an all-pole IIR filter to model this vocal-tract transfer function. The filter coefficients can be estimated from the segment of the speech signal by applying the Levinson–Durbin recursive algorithm.

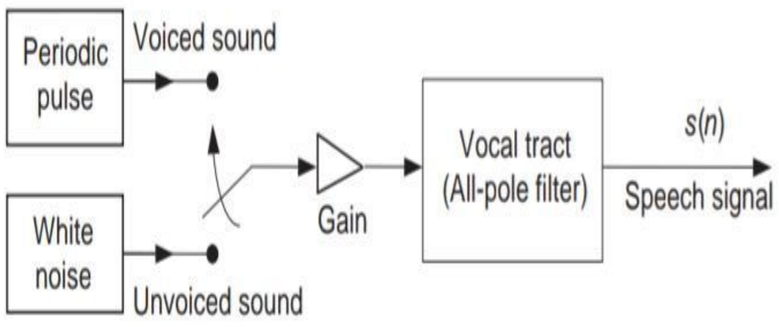

Figure :1 speech production model

To use the LPC model, the given segment of speech needs to be classified as either voiced or unvoiced. For the voiced signal, estimation of the fundamental frequency (also called pitch frequency) becomes very important. The inverse pitch frequency equals the pitch period, which is usually expressed in samples for the given sampling rate. The pitch period is used to generate the periodic pulse (also called pulse train) for exciting the LPC filter to produce the voiced speech. Several LPC-based speech codecs (coder–decoders), especially code-excited linear predictive (CELP) codecs at the bit rate of 8 kbps or lower, have been developed for wireless and network applications. CELP-type speech codecs are widely used in mobile and IP telephony communications, media streaming services, audio and video conferencing, and digital radio broadcasting.

2.2 CELP Coding

CELP algorithms are based on the LPC approach using the analysis-by-synthesis scheme. The coded parameters are analyzed to minimize the perceptually weighted error in the synthesized speech via a closed-loop optimization procedure. All CELP algorithms share the same basic functions including short-term synthesis filter, long-term pitch synthesis filter (or adaptive codebook), perceptual weighted error minimization procedure, and stochastic (or fixed) codebook excitation. The basic structure of the CELP algorithm is illustrated in Figure 2, where the top portion of the figure is the encoder and the bottom portion is the decoder.

In the encoder, the LPC and pitch analysis modules analyse speech to estimate the parameters for the speech synthesis model. They are followed by the speech synthesis module to minimize the weighted error. To develop an efficient search procedure, the number of operations can be reduced by placing the weighting filter.

In the decoder, the excitation signal e(n) is first passed through the long-term pitch synthesis filter P(z) and then the short-term LPC synthesis filter 1/A(z). The reconstructed signal ^xðnÞ is sent to the postfilter F(z) which emphasizes speech formants and attenuates the spectral valleys between formants.

2.2.1 Synthesis Filter

The time-varying short-term synthesis filter 1/A(z) is updated frame by frame using the Levinson–Durbin recursive algorithm. The synthesis filter 1/A(z) is expressed as

$$1/A(z) = \frac{1}{1-\sum_{i=0}^{p} aiz^{-i}}$$

where ai are the short-term LPC coefficients and p is the filter order. These coefficients are used to estimate the current speech sample from the previous samples. The widely used method to calculate the LPC coefficients is the autocorrelation method derived by minimizing the total squared errors between the speech samples and their estimates. A window (such as the Hamming window defined in Chapter 3) is applied to form a speech frame and calculate the autocorrelation coefficients based on the speech samples in the frame as

$$r_m(j) = \sum_{n=0}^{N-1-j} x\, x_m(n) x_m(n+j), \quad j = 0,1,2,\ldots\ldots ,p,$$

where N is the window (or frame) size, j is the autocorrelation coefficient index, m is the frame index, and n is the sample index inside the frame.

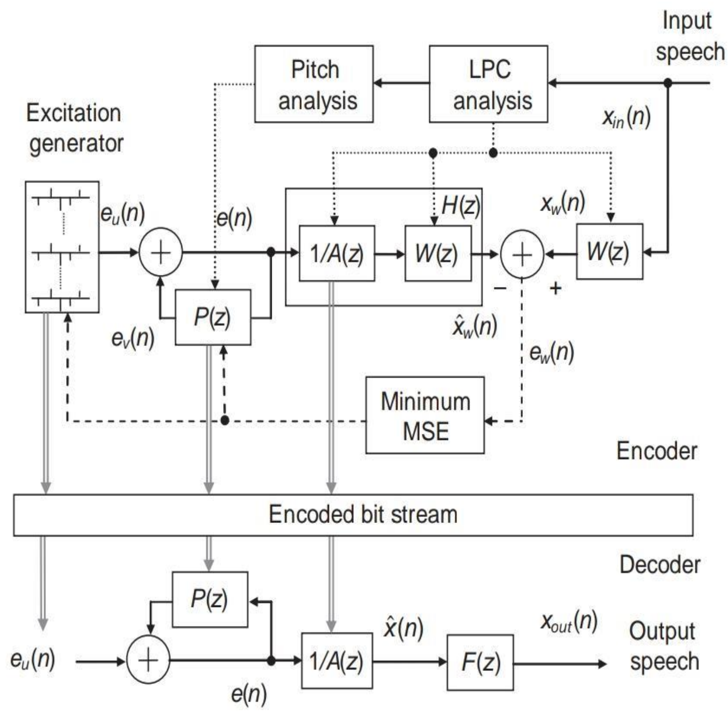

Figure 2: Block diagram of typical CELP algorithm

3. SPEECH ENHANCEMENT

In general, there are three different types of noise reduction techniques: single channel, dual channel, and multiple channel. The dual-channel technique is based on the adaptive noise cancellation introduced in Chapter 6. The multiple-channel methods can be realized as adaptive beamforming and blind source separation. In this section, we focus on single-channel techniques. There are three general classes of single-channel speech enhancement techniques: noise subtraction, harmonic-related noise suppression, and speech re-synthesis using vocoders.

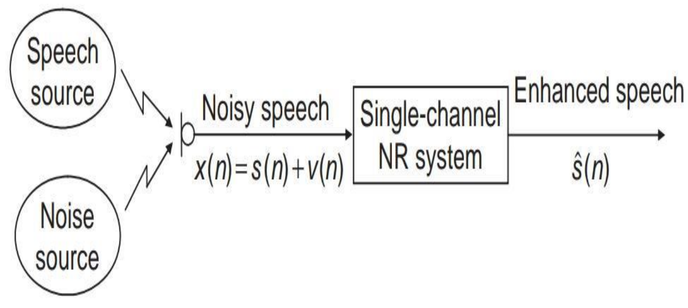

Fig.3. A Single channel speech enhancement system

The first technique suppresses noise by subtracting the estimated amplitude spectrum of the noise from the noisy signal, which will be discussed in following sections. The second method employs fundamental frequency tracking using adaptive comb filters for reducing periodic noise. The third technique focuses on estimating speech modeling parameters using iterative methods and uses these parameters to re-synthesize noise-free speech. The single-channel speech enhancement algorithms usually assume that the background noise is stationary (or quasi-stationary) and the characteristics of the noise can be estimated during the silent periods between utterances. Noise subtraction algorithms can be implemented either in the time domain or in the frequency domain. The frequency-domain implementation based on short-time magnitude spectra estimation for noise subtraction is called spectral subtraction. The spectral subtraction algorithm uses the DFT to obtain the short-time magnitude spectrum of the noisy speech, subtracts the estimated noise magnitude spectrum, reconstructs the DFT coefficients using the subtracted magnitude spectrum with the original phase spectrum, and performs the IDFT to obtain enhanced speech. In practical applications, the spectral subtraction technique uses the computationally efficient FFT and IFFT algorithms**.**

4. Real-Time Implementation Methodology

Real-time test has a great value, particularly for audio applications that have strict timing constraints such as audio streaming. In real-time application, the input signal and the generated output are processed continuously which explicates that the mean processing time per sample is lesser than the sampling period. Thus, we have tested our proposed approach on a flexible platform which is suitable with the particularity of our application. For this reason we have used a developed starter kit containing a DSK board based on DSK C6713 and the software tool (Code Compose studio). We have also used a rapid prototyping tool from Math-works.

4.1 DSK C6713 Overview

The DSK C6713 board incorporates a fixed point digital signal processor TMS320C6713 that works with a clock frequency of 1GHz. Furthermore, it is equipped with audio codec TLV320AIC23B (AIC23) which offers analog-to-digital conversion (ADC) and digital-to analog conversion (DAC) functions among a selecting sampling rate ranged of alternative settings from 8 to 96 kHz. As indicated in the figure 4 below the DSK board has four connections which provide analog inputs and outputs: A microphone input port, a line in port, a line out port, and a headphone port. The DSK board includes 16 MB (megabytes) of synchronous dynamic RAM (SDRAM) and 512 kB (kilobytes) of flash memory. The TMS320C6713 is based on the Very Long Instruction Word (VLIW)

architecture that is suitable for numerical exhaustive algorithms. The internal program memory is structured so that a total of eight instructions can be fetched every cycle. Figure 4 shows the Spectrum Digital DSK board.

Fig.4. TMS320C6713 DSK development board from Texas instruments

4.2 Rapid Prototyping Technology

For a rapid prototyping of real-time applications on DSP processor, we have employed the Embedded Target for C6000 DSPs Platform and RTW (Real-time Workshop) which convert the Simulink model into efficient C code especially for C6000 processors. MATLAB Link for CCS Development Tools interface with TI CCS (code composer studio) to generate an executable which is loaded into theDSK-C6713.

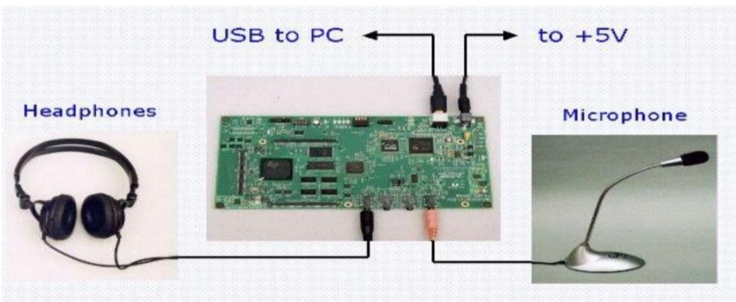

Fig.5. DSK C6713 Board

5. RESULTS

In this section, a MATLAB program has been developed to implement the speech enhancement system based on linear predictive coding techniques and adaptive synthesis filter. The database was taken from the TIMIT database 19, sampled at 16KHz and recorded by disturbed noise by male voice. For this purpose, the used mother wavelet is “db10”, the signal-to-noise ratio (SNR) ranging from 0-15db were used. To evaluate the efficiency of the developed algorithm and the proposed system is performed using SNR. The time evolution and spectrograms of the clean speech clean signals by LPC algorithm and our proposed approach when signal is corrupted by factory noise at SNR=10db. The proposed algorithm reduces interfering signals.

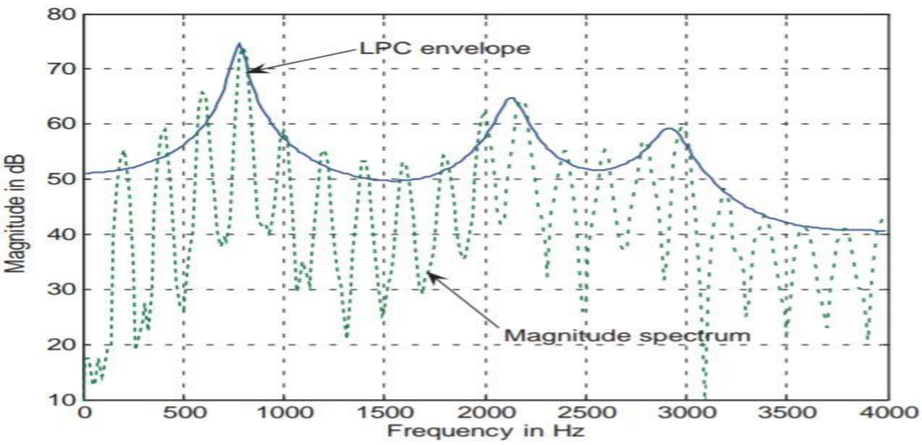

FIGURE 6: MAGNITUDE SPECTRUM OF SPEECH AND IT’S SPECTRAL ENVELOPE DERIVED FROM THE SYNTHESIS FILTER USING LPC COEFFICIENTS

6. CONCLUSION

In this paper, a new speech enhancement algorithm using Linear predictive coding techniques and adaptive synthesis filter has been presented. The proposed approach proves its reliability to improve the speech intelligibility without affecting the signal quality referring to the performance evaluation such as SNR. Finally, the real-time test of speech denoising has been successfully implemented in TMS320C6713 platform and reveals that the proposed algorithm has significantly improved the speech intelligibility**.**

References

1. ITU-T Recommendation (1996) G.723.1, Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 & 6.3 kbit/s, March.
2. CCITT Recommendation (1992) G.728, Coding of Speech at 16 kbit/s Using Low-delay Code Excited Linear Prediction.
3. ITU-T Recommendation (1995) G729, Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS- ACELP), December.
4. ITU-T Recommendation (2002) G722.2, Wideband Coding of Speech at Around 16 kbit/s Using Adaptive Multi-rate Wideband (AMR-WB), January.
5. 3G TS 26.190 V1.0.0 (2000-12) (2000) Mandatory Speech CODEC Speech Processing Functions; AMR Wideband Speech CODEC; Transcoding Functions (Release 4), December.
6. 3GPP TS 26.171 (2001) Universal Mobile Telecommunications System (UMTS); AMR Speech CODEC, Wideband; General Description (Release 5), March.
7. Bessette, B., Salami, R., Lefebvre, R. et al. (2002) The adaptive multi-rate wideband speech codec (AMR-WB). IEEE Trans. Speech Audio Process., 10 (8), 620–636.
8. Kondoz, A.M. (1995) Digital Speech Coding for Low Bit Rate Communications Systems, John Wiley & Sons, Inc., New York.
9. Kuo, S.M. and Morgan, D.R. (1996) Active Noise Control Systems – Algorithms and DSP Implementations, John Wiley & Sons, Inc., New York.
10. Tian, W., Wong, W.C., and Tsao, C. (1997) Low-delay subband CELP coding of wideband speech. IEE Proc. Vision Image Signal Process., 144, 313–316.