LPC Models and Different Speech Enhancement Techniques- A Review

Mr. Arjun Ghule
Research Scholar, Department of ECE, J.N.T.U., Hyderabad, India
Dr.Prabhu Benakop
Director, Aurora`s Engineering College, Bongir, Hyderabad, India

Abstract

Author has already published one review paper on the quality enhancement of a speech signal by minimizing the noise. This is a second paper of same series. In last two decades the researchers have taken continuous efforts to reduce the noise signal from the speech signal. This paper comments on, various study carried out and analysis proposals of the researchers for enhancement of the quality of speech signal. Various models, coding, speech quality improvement methods, speaker dependent codebooks, autocorrelation subtraction, speech restoration, producing speech at low bit rates, compression and enhancement are the various aspects of speech enhancement. We have presented the review of all above mentioned technologies in this paper and also willing to examine few of the techniques in order to analyze the factors affecting them in upcoming paper of the series.

Key words: LPC models, coding, speaker dependent codebooks, autocorrelation subtraction, speech restoration, compression and enhancement of speech

Introduction

In last two decades, various models have been developed by the researchers for speech enhancement. Many people have worked for analysis of the models and few others have commented on the advantages and disadvantages of the models. We are writing the review of some models developed and principles used while developing them. The source filter model of speech is widely used for digital coding of speech; the popularity of the LPC method stems from its computational efficiency in determining the synthesis filter parameters [2]. Speech enhancement algorithms that reduce the level of noise present in a speech signal captured in the presence of background noise have several applications [5].

Literature Review

Evangelos E. Milios and Alan V. Oppenheim [1] The work carried out by various researchers has demonstrated the importance of the long-time phase of speech. The possibility of using the long-time phase of the LPC residual signal in speech synthesis is investigated by author.

The importance of the various parts of the long-time spectrum of the residual is also suggested. The potential for synthesizing high-quality speech using the phase-only residual has been demonstrated. It is recommended that coding the long-time phase together with either the smoothed long-time magnitude alone or the smoothed long-time magnitude augmented by some more accurate low frequency magnitude information.

V. K. Jain, R. Hangartner [2] The introduction of multipulse excitation for LPC coders has increased the quality achievable for digitally coded speech at bit rates in the 9.6 kbs range. A simplified multipulse analysis is proposed with particular prominence on the speech model developed as a result of a two pass method. In the first pass, estimated LPC parameters generated by conventional covariance analysis are used to generate the forward prediction error; the multi—pulse sequence is them detected by thresholding the residual. The second pass generates the final LPC parameters by a covariance analysis incorporating knowledge of the estimated pulse locations and amplitudes along with perceptual synthesis error weighting. Experimental results are presented to demonstrate the method.

Rizwan Ishaq and Begofia Garcia Zapirain [3] The vocal fold modulates the air source from lungs to produce voicing source for speech production. The vocal fold essential part of speech production resides in larynx. The larynx cancer treatment necessitates removal of larynx, in consequences normal speech production destroyed due to no voicing source available. The new voicing source provided artificially or by use of Paryngo-esophageal (PE) segments. The voicing source or residual signal for Esophageal (E) speech uses PE segment, has irregular behavior, and produces degraded quality speech. This paper discussed and evaluated the residual signal or voicing source enhancement of E speech by incorporating speech enhancement method Adaptive Gain Equalizer (AGE), in time frequency and modulation frequency along with formants modification by Line Spectral Frequencies (LSF) and Linear Predicative Coding (LPC). The system validated by measuring Harmonic to Noise Ratio (HNR) temporally and maximum of 4dB enhancement has been observed in comparison to Kalman filtering based enhancement where enhancement observed maximum of 2dB.
The improved HNR has validated the system capability to improve quality of E speech. The system successfully removed noise, as well tried to enhanced residual signal for better and intelligible E speech signal. The MAGE modulation frequency system provides better enhancement in comparison to AGE and overall both systems outperformed Kalman filtering based system [4]. Along with voicing source enhancement, formant enhancement significantly improves quality of E speech in comparison to poles modification by shifting upward [4] when conducted listening test. Although LPC analysis/synthesis provides good estimation of voicing source signal, but as comparison to normal speech this estimation in not perfect and modeling of PE segment can provides better voicing source signal which can be improved and modified through this system. The future can be to provide optimized value of L opt by having noise information from modeling of PE segment.

D. Hanumantha Rao Naidu, G. V. Prabhakara Rao, Sriram Srinivasan [5] Speech enhancement techniques that employ trained codebooks of speech and noise linear predictive coefficients have been previously shown to provide good performance in non stationary noise environments. The speech codebooks are typically trained using data from a large number of speakers. For noise reduction systems in personal devices such as mobile phones, the use of speaker dependent codebooks is appealing as such devices are typically owned by a single user.

It is however not straightforward whether the use of speaker dependent codebooks translates into an observable benefit in terms of noise reduction. Moreover, it is of interest to quantify the improvement obtained over speaker independent codebooks. In this paper, the benefit of using speaker dependent codebooks is analyzed in terms of the resulting improvement in noise reduction when compared to using speaker independent codebooks.

The hypothesis that the use of speaker dependent (SD) codebooks can provide improved performance compared to speaker independent (SI) codebooks when used for speech enhancement has been validated through experimental results. It has been seen that the ability of SD codebooks to better model the spectral shapes corresponding to a particular speaker's data also translates into improvements in noise reduction using a codebook-based speech enhancement method. SD codebooks provide a significant improvement over SI codebooks in terms of all three performance measures considered in this paper: segmental signal-to-noise ratio, log spectral distortion and PESQ scores, and indicate that the benefits justify the added complexity in training the SD codebooks on-line. In fact, depending on the input SNR, it was seen that an SD codebook of half the size of an SI codebook is sufficient to provide a similar level of performance. Reducing the codebook size can significantly reduce the computational complexity of the speech enhancement algorithm. This is desirable as the resulting savings more than compensate for the added complexity in adapting the speech codebooks to the user's voice, which only needs to be done when the user changes. Our results indicate that speaker dependent approaches are promising for applications such as speech enhancement for personal devices, e.g., mobile phones. Future work will extend the investigation to gender and language dependent codebooks. Another interesting extension is when the number of users is larger than one, but limited to a small number. Examples of such applications include VoIP calls on a PC or home telephony systems, where the users could be the members of a family. Future work will investigate the use of one speech codebook per person, where the system automatically determines the appropriate codebook to be used based on the input signal.

C. K. Mn and K. Y. Choi [6] A robust linear predictive coding (LPC) method that can be used in noisy as well as quiet environment has been studied. In this method, noise autocorrelation coefficients are first obtained and updated during non speech periods. Then, the effect of additive noise in the input speech is removed by subtracting values of the noise auto correlation coefficients from those of autocorrelation coefficients of corrupted speech in the course of computation of linear prediction coefficients. When signal-to-noise ratio of the input speech ranges from 0 to 10 dB, a performance improvement of about 5 dB can be gained by using this method. The proposed method is computationally very efficient and requires a small storage area. Authors have studied linear predictive coding in noisy environment and proposed a method to reduced gradation caused by additive white noise. The approach is based on subtraction of autocorrelation coefficients of noise from those of corrupted speech after estimation of noise periodogram during intervals of non-speech activity. By using the proposed method, one can improve the performance of an LPC vo coder by about 5 dB in SNR.

A.B. Premkumar, Ang Eeluang and A. S. Madhukumar[7] Restoration of speech parameters that have been degraded by noisy channel is of paramount importance, since mutilated speech parameters such as Linear Predictive Coefficients may render speech totally irrecoverable. The objective of this paper is to provide a technique using discrete dyadic wavelet transforms as a preprocessing stage prior to LPC extraction. We show that this step makes the speech parameters more robust to noise perturbation in the channel and hence makes recovery of speech possible. We discuss the effects of applying a single level and a three level transformation to speech prior to compression and compare the quality of recovered speech in both cases. We also apply Wiener filter to the wavelet coefficients and show that the recovered speech is of appreciable quality.

Speech recovery from channel noise has been studied. Recovery has been made possible using DWT techniques. Applying DWT analysis prior to the extraction of LPC parameters and synthesizing speech from the inverse DWT coefficients has proved effective in combating channel noise. This method works well so long as the formant information in speech is not completely destroyed by noise affecting the LPC coefficients

Siva Sushma G, D R Sandeep [8] Over the past several years there has been considerable attention focused on compression and enhancement of speech signals. This interest is progressed towards the development of new techniques capable of producing good quality speech at the output. Speech compression is a process of converting human speech into efficient encoded representations that can be decoded to produce a close approximation of the original signal. In this paper the simulated vo coder (LPC) using mat lab was implemented for compression. The result obtained from LPC was compared with other implemented speech compression using Wavelet transform in terms of quality. Speech enhancement means improving the quality or value of the signal in a noisy environment. In this paper we proposed spectral subtraction (S.S) and wavelet transform methods for de-noising and the result of one method was compared with other. From the results we see that in both compression and enhancement, the performance of wavelet transform was better than other.

In this paper, speech compression using LPC, Wavelets and enhancement using S.S, wavelet transforms was presented. In both compression and enhancement of speech signals using wavelet transforms, we have chosen DB20 wavelet and then the performance was tested and the following points were observed. From the output signals of Lpc, we can observe that the noise in the synthesized files is stronger than in the actual signals, but in wavelets the noisy feeling is low. By this we can say that the signal to noise ratio (as one of the most important measurements of the performance) of the Wavelet transform has high values comparable to LPC and high compression is achieved using wavelet transforms. From the enhancement results, we can observe that the signal to noise ratio of spectral subtraction is low when compared to wavelets. By this we can say, that the spectral subtraction works poorly in de-noising when compared to the wavelets. The analysis that we undertook for wavelets in both compression and enhancement includes only the compactly supported wavelets. An improvement in the performance can be done by studying the effects of other wavelets on compression and enhancement.

Bishnu S. Atal and Joel R. Remde [9] The excitation for LPC speech synthesis usually consists of two separate signals - a delta-function pulse once every pitch period for voiced speech and white noise for unvoiced speech. This manner of representing excitation requires that speech segments be classified accurately into voiced and unvoiced categories and the pitch period of voiced segments be known. it is now well recognized that such a rigid idealization of the vocal excitation Li often responsible for the unnatural quality associated with synthesized speech. This paper describes a new approach to the excitation problem that does not require a priori knowledge of either the voiced-unvoiced decision or the pitch period. All classes of sounds are generated by exciting the LPC filter with a sequence of pulses; the amplitudes and locations of the pulses are determined using a non-iterative analysis-by-synthesis procedure. This procedure minimizes a perceptual-distance metric representing subjectively-important differences between the waveforms of the original and the synthetic speech signals. The distance metric takes account of the finite-frequency resolution as well as the ditTerential sensitivity of the human ear to errors in the formant and inter-formant regions of the speech spectrum.

Our work on developing a new model of LPC excitation for producing high quality speech is as yet very preliminary. We find that speech quality can be significantly improved by using multi pulse excitation signal. Moreover, the locations and amplitudes of the pulses in the multi-pulse excitation can be determined by using a computationally efficient non-iterative analysis-by synthesis procedure that can minimize the perceptual difference between the natural and the synthetic speech waveforms. The difficult problems of voiced-unvoiced decision and pitch analysis are eliminated.

Conclusion

Author has already published one review paper on the quality enhancement of a speech signal by minimizing the noise. This is a second paper of same series. This paper comments on various models and methodologies of the speech enhancement techniques. The various models proposed by the scholars were working on various factors and the researchers also commented on the methods for improving the performance of the models. All the models are good for various surrounding conditions and there are several factors affecting the performance of each model. We have presented the review of various mentioned technologies in this paper and also willing to examine few of the techniques in order to analyze the factors affecting them in upcoming paper of the series.

References

1] *“The Phase-only Version of the LPC Residual in Speech Coding*”- Evangelos E. Milios and Alan V. Oppenheim, ICASSP 83, BOSTON

2] *“Efficient Algorithm for multi—pulse LPC analysis of speech*”, V. K. Jain, R. Hangartner, ICASSP— 84

3] *“Adaptive Gain Equalizer for Improvement of Esophageal Speech”,* Rizwan Ishaq and Begofia Garcia Zapirain, 978-1-4673-5604-6/12 ©20 12 IEEE.

4] B. Garcia and A. Mendez, "*Oesophageal speech enhancement using poles stablization and kalman filtering,"* ICASSP, pp. 1597-1600, 2008.

5] *“Speech Enhancement Using Speaker Dependent Codebooks”*, D. Hanumantha Rao Naidu, G. V. Prabhakara Rao, Sriram Srinivasan, 978-1-4577-0274-7/11/©2011 IEEE, DSP 2011.

6] “Improving Lpc Nalysis Of Noisy Speech By Autocorrelation Subtraction Method”, C. K. Mn and K. Y. Choi , CH161O-5/81/0000-1082 © 1981 IEEE

7] “Speech Restoration From Lpc Parameters Corrupted By Channel Noise”, A. B. Premkumar, Ang Eeluang and A. S. Madhukumar, 1999 IEEE TENCON

8] *“Compression and Enhancement of Speech Signals”*, Siva Sushma G, D R Sandeep, Chennai and Dr.MGR University Second International Conference on Sustainable Energy and Intelligent System (SEISCON 2011) , Dr. M.G.R. University, Maduravoyal, Chennai, Tamil Nadu, India. July. 20-22, 2011.

9] “A New Model Of Lpc Excitation For Producing Natural-Sounding Speech At Low Bit Rates”, Bishnu S. Atal and Joel R. Remde, CH 1746-718210000 . 0614 © 1982 IEEE