Characterization of Speech Signal in LabVIEW Platform

Amrutha .S[1], Athira S Nair[1], Catherine J Mathew[1], Jeena Elsa George[1], Mr. Mathew George[2]
[1] UG scholar, [2]Assistant Proffessor
Department of Electronics and Communication Engineering
Amal Jyothi college of Engineering, Kanjirapally, Kottayam, India

Abstract – **Speech has remained the most desirable medium of communication between humans. Speech signal is a highly-correlated signal which possesses both short- and long-term similarities. These similarities or redundancies can easily be modelled by very compact LPC and pitch filter formulations. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters. We establish this technique for estimating basic speech parameters pitch and formants. This paper presents a visualization tool for speech processing aspects based on National Instruments LabVIEW trade environment**.
Index term: **formant, LPC, pitch**

INTRODUCTION

Speech is an immensely information-rich signal exploiting frequency-modulated, amplitude-modulated and time-modulated carriers (e.g. resonance movements, harmonics and noise, pitch intonation ,power, duration) to convey information about words, speaker identity, accent ,expression, style of speech, emotion and the state of health of the speaker. All this information is conveyed primarily within the traditional telephone bandwidth of 4 kHz. The speech energy above 4 kHz mostly conveys audio quality and sensation.

Speech processing is the study of speech signals and the processing methods of these signals. The purpose of processing speech signals is to enhance and extract information, which is helpful in providing as much knowledge as possible about the signal's structure i.e., about the way in which information is encoded in the signal. Speech processing is the study of speech signals and the processing methods of these signals. Aspects of speech processing includes the acquisition, manipulation, storage, transfer and output of speech signals.The need for more powerful signal processing methods and more flexible design strategy is exponentially increasing in many applications. Increase in complexity of the problem so as the solution, often times demands development of sophisticated signal processing algorithms, thereby, implying higher system cost. The key system design strategy remains to be achieving maximum performance and flexibility per cost.

We describes a very compact LPC and pitch filter formulations technique to estimate basic speech parameters formants and pitch. This paper presents a visualization tool for speech processing aspects based on National Instruments LabVIEW trade environment.

LPC MODEL

Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envolpe of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.
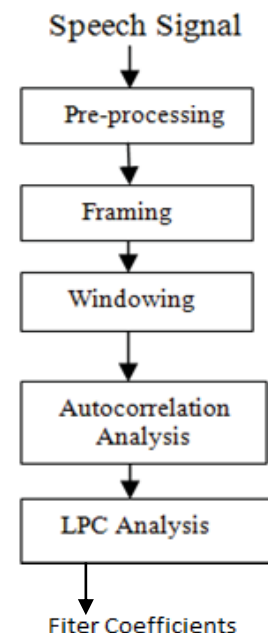

Fig 1:Block diagram of LPC model

In our work we started with linear prediction coefficients (LPC) as a basic feature. In course of extracting the features of speech signal, first we pre-processed the speech signal where the signal was passed

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCETET-2015 Conference Proceedings

through a first order pre-emphasis filter to spectrally flatten the signal and to make it less susceptible to finite precision effects later in the signal processing, then normalization, mean subtraction and silence removal was performed. After preprocessing the speech signal was divided into frames of 30ms with an overlap of 20ms to avoid the effect of discontinuity. The speech signal was sampled at 16 KHz. Then each frame is passed through a hamming window. For each frame, LPC are calculated. The basic idea behind the LPC model is that a given speech sample sample $s(n)$ at time n, can be approximated as a linear combination of the past p speech samples such that

$$S(n) = a_1 s(n-1) + a_2 s(n-2) + \ldots + a_p S(n-p) \quad (1)$$

Where the coefficients a1, a2, ape are assumed constant over the speech analysis frame. Equation (1) is converted to equality by including an excitation term G u (n) giving ,

$$S(n) = \sum_{k=1}^{p} a_k s(n-k) + Gu(n) \quad (2)$$

where u(n) is a normalized excitation, G is the gain of excitation and p is the order of LPC analysis. By expressing output signal s (n) and input signal u (n) in z-domain, an all-pole system with transfer function is shown in equation

$$H(z) = \frac{S(Z)}{GU(Z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k Z^{-1}} = \frac{1}{A(Z)} \quad (3)$$

The LPC order p typically lies in the range of 8 to 20. In this paper the order of LPC is taken sixteen. After this the filter coefficients (corresponding to the vocal tract) are derived by minimizing the mean square error between the input s(n) and the estimated sample s(n) as shown in equation

$$\acute{s}(n) = \sum_{k=1}^{p} a_k \, s(n-k) \quad (4)$$

$$e(n) = s(n) - \acute{s}(n) \quad (5)$$
$$E = \sum_{n=-\infty}^{\infty} e^2(n) \quad (6)$$

The basic idea of LPC is to transmit the prediction error (residue) instead of the speech signal. Since a linear predictor with properly chosen order can predict the signal with relatively small error variance, the power of residuals is effectively smaller than the power of the original signal. This property of linear prediction enables us to use lower bit rate for transmitting the speech signal through a communication channel.

FORMANT DETECTION

Formants are defined by the spectral peaks of the sound spectrum of the voice. In speech s Formants are the resonant frequencies of vocal tract and formant frequencies vary with a vocal tract configuration. Typically there are three resonance of significance for human vocal tract that is typically below 3500 Hz. The phonemes can be easily distinguished by the frequency values of the first two or three formants, which are called as F1, F2 and F3. F1 varies from 300 Hz to 1000 Hz, F2 varies from 850 Hz to 2500 Hz and F3 can vary from 2300 Hz to 3500 Hz. The

energy concentration of formant frequency is more than any other frequency of speech signal.

Using LabVIEW to Detect Formants

Several methods can be used to detect formant tracks . The most popular method however is the Linear Prediction Coding (LPC) method. This method applies an all-pole model to simulate the vocal tract.
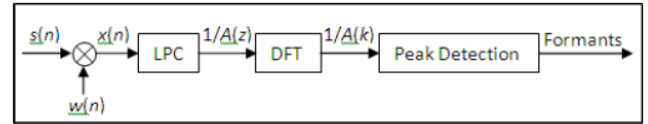

Fig 2: Formant detection with LPC method

Applying the window w(n) breaks the source signal s(n) into signal blocks x(n). Each signal block x(n) estimates the coefficients of an all-pole vocal tract model by using the LPC method. After calculating the discrete Fourier transform (DFT) on the coefficients A(z), the peak detection of 1/A(k) produces the formants.
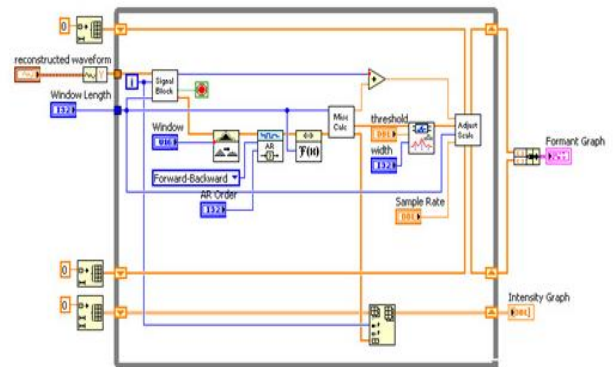

Fig4:Formant Generation Sub VI

PITCH DETECTION

The term "Pitch" corresponds to the name given to the fundamental frequency of a speech signal. This value can be easily seen within the semi-stationary speech waveform signal as the time interval from one peak to the next. The periodic opening and closing of the vocal folds results in the harmonic structure in voiced speech signals. The inverse of the period is the fundamental frequency of speech. Pitch is the sensation of the fundamental frequency of the pulses of airflow from the glottal folds. The terms pitch and fundamental frequency of speech are used interchangeably. The pitch of the voice is determined by four main factors.

Using LabVIEW to Detect Pitch

Figure 5 shows the flow chart of pitch detection with the LPC method. This method[6][7] uses inverse filtering to separate the excitation signal from the vocal tract and uses the real cepstrum signal to detect the pitch.
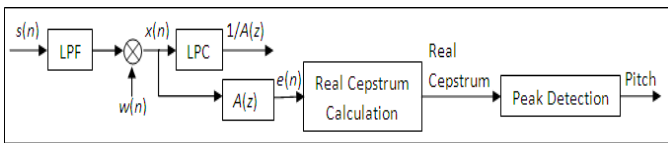
Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCETET-2015 Conference Proceedings

Fig 3: Pitch Detection with the LPC Method

In Figure 5, the source signal *s(n)* first goes through a low pass filter (LPF), and then breaks into signal blocks *x(n)* by applying a window *w(n)*. Each signal block *x(n)* estimates the coefficients of an all-pole vocal tract model by using the LPC method. These coefficients inversely filter *x(n)*. The resulting residual signal *e(n)* passes through a system which calculates the real cepstrum. Finally, the peaks of the real cepstrum calculate the pitch.
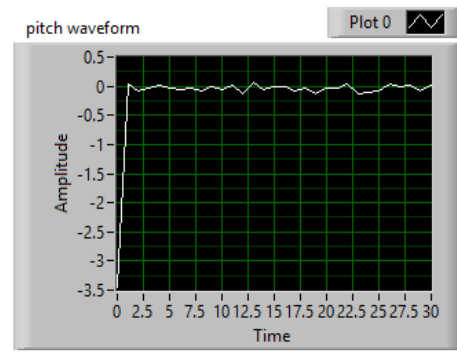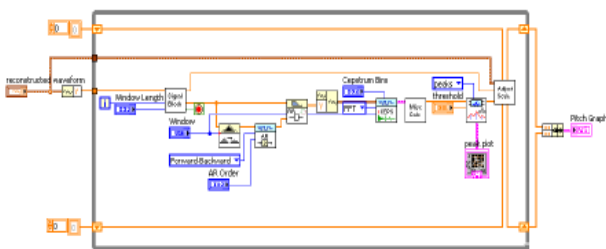

Fig5: Pitch Generation Sub VI

EXERIMENTAL RESULT

We carried out the proposed method for the estimation of formant and pitch.The graph of the estimated formant and the pitch of the sound file is as shown in the figure.
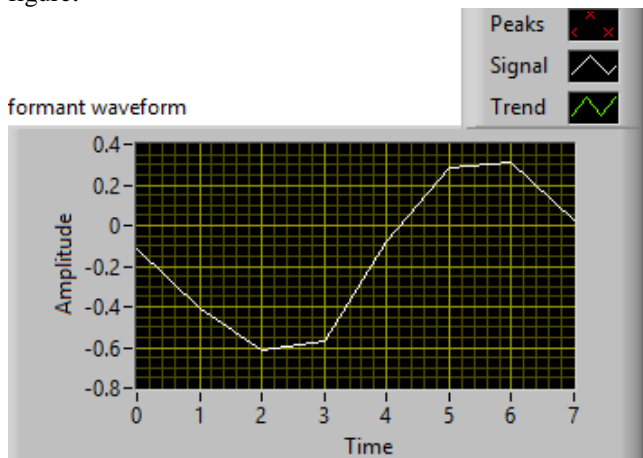

Fig 6:Formant waveform

Fig 7: Pitch waveform

CONCLUSION

In this paper, we proposed a method to estimate the fundamental speech parameters pitch and formants. Here we used LPC and cepstral analysis .Considering the efficiency of the results obtained, it is concluded that the algorithm implemented in LabView is working successfully.

REFERENCE

[1] Sumit Srivastava, *"Formant based linear prediction coefficients for speaker identification"* International Conference on Signal Processing and Integrated Networks (SPIN), 2014.

[2] Md Fozur Rahman Chowdhury*, "Text independent distributed speaker identification and verification using GMM UBM speaker models for mobile communications,"* 10th International Conference on Information Science, Signal Processing and Their Application, 2010, pp 57-60.

[3] Tomi Kinnunen, Evgeny Karpov and Pasi Franti (2006) *"Real- time speaker identification and verification"*, IEEE Transaction on Audio, Speech and Language Processing, Vol. 14, No.1, pp. 277- 278.

[4] L.R.Rabiner and B.H.Juang, *Fundamentals of Speech Recognition*, 1st ed., Pearson Education, Delhi, 2003.

[5] J. Makhoul, *"Linear prediction: A tutorial review,"* Proc. of IEEE, vol. 63, no. 4, pp. 561-580, 1975.

[6] R.C.Snell and F. Milinazzo, *"Formant location from LPC Analysis data,"* IEEE Transactions on Speech and Audio Processing, vol. 1, no. 2, pp. 129–134, Apr. 1993.