Comparison Between Different Methods for Epoch Extraction from Speech Signal

B. V. Pathak
Research scholar
Priyadarshini College of Engineering, Nagpur

A. R. Panat
Research Guide
Priyadarshini College of Engineering, Nagpur

Shreya. R. Garipalli
MKSSS's Cummins College of Engineering, Pune

Abstract

Speech is produced by exciting the vocal tract system by an excitation source. The excitation is significant at the instants of glottal closure. These instants of significant excitation are called Epochs. Epochs play a significant role in speech processing and is used are used in many applications. There are different methods for extraction of epochs from speech signal. In this paper we have discussed and compared three methods namely Linear Prediction Based (LP Based) Method, Group Delay Based (GD Based) Method and Zero Frequency Filtering (ZFF) Method. LP-Based method and GD-Based method depends on characteristics of both the excitation source and vocal tract system. ZFF method is based on the impulse nature of the excitation signal and is not affected by the characteristics of the vocal tract system. The ZFF method gives more accurate and reliable results compared to the first two methods.

Keywords: Epoch, Glottal Closure, Linear Prediction, Group Delay, Zero Frequency Filtering

I. INTRODUCTION

The excitation to the vocal tract system is responsible for producing the speech signal. This excitation may be due to: 1. Glottal vibration, 2. Burst, 3. Frication [1]. For voiced speech the excitation to the vocal tract system is glottal vibration. Glottal vibration is nothing but the opening and closing of the glottis i.e. vibration of the vocal cords. The excitation to the vocal tract system is present throughout the speech production, but it is significant at instants of glottal closure. These instants of glottal closure are known as Epochs. The excitation at epochs is significant because the energy at that instant is large compared to energy at neighboring instants i.e. excitation is Impulse like. Due to this property the excitation signal is approximated as sequence of impulses with different amplitudes i.e. strengths.

Extraction of epochs is important in Speech Analysis. If the information of the epochs is available then we can determine the characteristics of excitation source and vocal tract system like fundamental frequencies and formant frequencies. Glottal activity detection is also possible with the knowledge of epochs which is then used for the separation of the voiced and unvoiced parts of the speech signal [3]. After the extraction of epochs we can use them as pitch markers and modify the epoch sequence according to the desired signal i.e. speech transformation [7]. Epochs are also useful for multispeaker separation. In multispeaker separation the speech components of every speaker are separated from the mixture of speech signals collected using the recording device which is placed in multispeaker environment [8].

Electroglottography (EGG) method can be used for epoch extraction. In this method the EGG signal is obtained by placing the electrodes across the neck. The EGG signal measures the laryngeal impedance. Then the time derivative of the EGG signal is computed and epochs occur at the maximum of the time derivative of EGG signal. The negative peaks of the differenced EGG signal denotes epochs [2].

Epochs can also be extracted using the LP based method. In this method the LP residual is computed using the LP analysis. The LP residual signal has maximum values at the instants of significant excitation but epoch detection is not done directly from LP residual. The accurate detection of epochs from LP residual is not possible because of the presence of multiple peaks of either polarity. Epochs can be detected unambiguously by computing the Hilbert Envelope (HE) of LP residual. The peaks in the HE of LP residual are known as instants of significant excitation [4].

The GD Based method for epoch extraction is based on the properties of the minimum phase signals and group delay function. LP residual of the speech signal is always a minimum phase signal. The group delay function of the LP residual is computed and the negative derivative of the group delay function gives phase slope value. The phase slope function is obtained by computing the phase slope value at every time instant. The positive zero crossings of the phase slope function are referred as epochs [5].

In this paper the LP based method and GD Based method for epoch extraction are discussed and the ZFF method is implemented. The ZFF method is based on the assumption that the excitation signal exhibits the impulse like characteristics and it is approximated as sequence of the impulses. The signal is then passed through two zero frequency resonators and the output is ZFF signal. The positive to negative zero crossings of the ZFF signal represents the epochs. Advantage of using the ZFF method is that the block processing approach is not used and the LP parameters are not required to compute the LP residual [1].

All rights reserved by *www.ijirst.org*

II. GROUP DELAY BASED METHOD FOR EPOCH EXTRACTION

Epochs in GD based method are extracted using the properties of the minimum phase signal and the group delay function. Epochs can be directly extracted from the speech signal by computing the average slope of the unwrapped phase spectrum of the signal. When we take the DFT of the speech signal and compute the phase spectrum of the signal, there is phase wrapping i.e. the phase values are present in the range of ($-\pi$ to $+\pi$) [5]. In order to compute the average slope we need to first determine the unwrapped phase. The process of phase unwrapping is difficult. To extract epochs we need to find only the average slope of the phase spectrum, there are many ways to compute this average slope. One method is to compute the average slope directly from the group delay function. In case of minimum phase signals the average group delay is zero. The speech signals may not be always minimum phase signals but the LP residual signal is always a minimum phase signal. The LP residual signal is obtained by inverse filtering in the Linear Prediction analysis. The average slope of the minimum phase signal represents the instants of significant excitation. Therefore group delay is computed always from LP residual [6].

Steps for finding the epochs using the GD based method are given below:

1) LP residual of the speech is computed using the Linear Prediction analysis of 10^{th} order, 20ms frame with shift of 10ms.
2) Then for each 10ms segment of LP residual, the group delay function is calculated.

The group delay function is calculated using the following formula:

$$\tau(\omega) = \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{X_R^2(\omega) + X_I^2(\omega)}$$

where, $X(\omega)$ and $Y(\omega)$ are the DFT of the windowed LP residual signal x(n) and nx(n), X_R and X_I are the real and imaginary parts of $X(\omega)$, Y_R and Y_I are the imaginary parts of $Y(\omega)$.

3) The discontinuities in the group delay function are removed by smoothing the group delay function.
4) Then the phase slope value is calculated by taking negative derivative of the smoothed group delay function.
5) Calculate the phase slope value for every time instant to get phase slope function.
6) The positive zero crossings of the phase slope function represents the instants of significant excitation.

III. LINEAR PREDICTION BASED METHOD FOR EPOCH EXTRACTION

Instants of significant excitation can be extracted using the Linear Prediction based method. The GD based method discussed earlier is computationally complex since the group delay is computed at every sample. In LP based method, we need to find the approximate locations of the instants of significant excitation then compute the group delay for only few samples around the approximate locations of instants of significant excitation. The LP residual of the speech signal is computed first. The LP residual is computed by inverse filtering in the linear prediction analysis. We can find the future sample using the linear weighted combination of the previous samples in forward linear prediction. The difference between the actual value and approximated value is the error signal i.e. LP residual [4, 6].

At the instants of significant excitation the value of the LP residual is large but direct detection of epochs using the LP residual results ambiguous epochs. Reason behind not using LP residual for epoch detection is it contains multiple peaks of either polarity. In order to detect instants of significant excitation unambiguously we need to compute the Hilbert Envelope of the LP residual signal. Hilbert Envelope of LP residual is a positive function. The HE of LP residual has peaks at the instants of significant excitation i.e. Epochs.

Steps to extract epochs from speech signal using the linear prediction method are:

1) LP residual is computed from the speech signal using the LP analysis.
2) The Hilbert Envelope of the LP residual is computed as.

$$h_e(n) = \sqrt{e^2(n) + e_h^2(n)}$$

where, $e_h(n)$ is the IDFT of the $E_h(k)$

$$E_h(k) = -jE(k) \quad \text{for k = o to (N/2)-1}$$
$$= jE(k) \quad \text{for k = (N/2) to (N-1)}$$

where, e(n) is the LP residual signal and E(k) is the DFT of the LP residual signal.

3) The Glottal Closure Evidence plot is required to find the approximate instants of significant excitation. The glottal closure evidence plot is computed by convolving the Hilbert Envelope with Gabor Function [6].

The Gabor function is given as

$$g(n) = \frac{1}{\sqrt{2\pi}\sigma} e^{(-(n-N/2)^2/2\sigma^2 + j\omega n)}$$

where, σ is the spatial spread of the Gaussian, ω is the frequency of sinusoid, n is the time index varying from 1 to N, N and is the length of the filter.

4) The positive zero crossings of the glottal closure evidence plot gives the approximate locations of the epochs.

All rights reserved by *www.ijirst.org*

5) The phase slope function of the LP residual is computed for the samples around approximate instants of significant excitation.
6) The positive zero crossings of the phase slope function gives the accurate locations of Epochs.

IV. ZERO FREQUENCY FILTERING METHOD FOR EPOCH EXTRACTION

In case of voiced speech the excitation to the vocal tract system is glottal vibration. The excitation is significant only at the instant of glottal closure. The discontinuities in the excitation signal are present due to the sharp closure of the glottis i.e. vocal folds. These discontinuities in the excitation signal can be represented by the sequence of impulses of various amplitudes. It is difficult to extract the instants of significant excitation directly from the speech signal. In order to extract epoch accurately speech signal is passed through the narrowband filter. Impulse in time domain represents all frequencies in frequency domain. The narrowband filter selected in this method is zero frequency resonator. The reason for selecting zero frequency resonator is that the effect of impulse is reflected even across zero frequency and vocal tract has free resonances at much higher frequency than zero frequency. In ZFF method two zero frequency resonators are used to reduce the effects of high frequency resonances [1, 2].

Steps for epoch extraction using ZFF method are given below:

1) The differenced signal x(n) is obtained from speech signal to reduce the slowly varying components in the speech signal due to recording devices.

$$x(n) = s(n) - s(n-1)$$

2) The differenced signal x(n) is passed through first zero frequency resonator.

$$y_1(n) = -\sum_{k=1}^{2} a_k y_1(n-k) + x(n)$$

where, $a_1 = -2$; $a_2 = 1$

3) The output of first zero frequency resonator is given as input to the second zero frequency resonator.

$$y_2(n) = -\sum_{k=1}^{2} a_k y_2(n-k) + y_1(n)$$

where, $a_1 = -2$; $a_2 = 1$

Passing the differenced signal through one zero frequency is equivalent to integrating it two times here we are using two resonators i.e. we are integrating the signal four times.

Figure 1 shows the plots of the speech signal and the differenced signal of corresponding speech signal. Figure 2 shows the plots of the outputs of the first and second zero frequency resonators respectively.

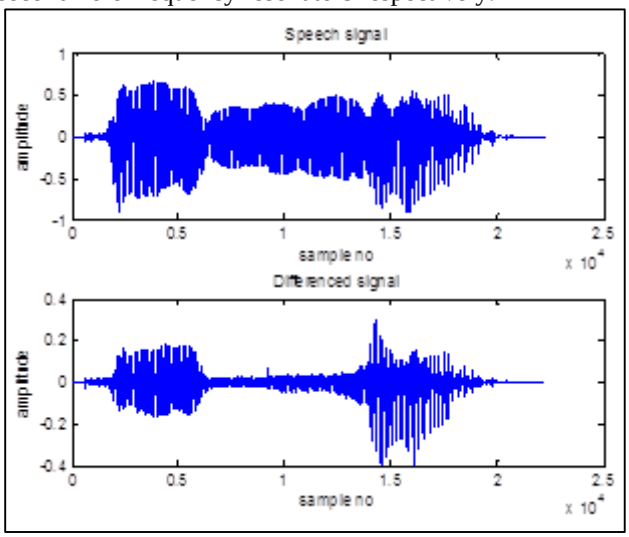


Fig. 1: Speech signal and differenced speech signal

All rights reserved by *www.ijirst.org*

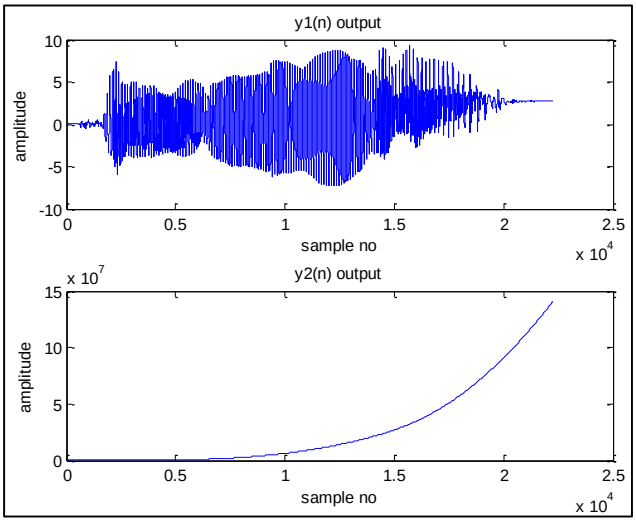


Fig. 2: Output of first and second zero frequency resonator

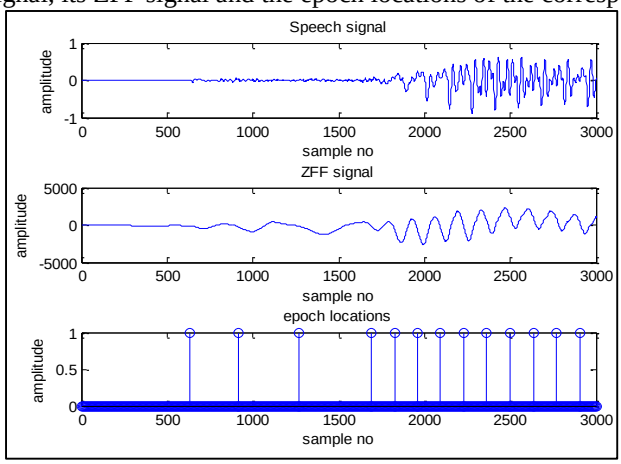
4) The average pitch period of the difference signal over 30ms segment is calculated using auto-correlation method.
5) The local mean of the y2(n) is calculated over twice the pitch period and is subtracted from output of the second zero frequency resonator to obtain Zero Frequency Filtered (ZFF) signal.

$$y(n) = y_2(n) - \frac{1}{2N+1}\sum_{m=-N}^{+N} y_2(n+m)$$

where, (2N+1) in the above equation corresponds to the number of samples in the window of twice the pitch period. The width of this window should in the range of 5-10ms. In this paper 10ms window is used.

6) The positive to negative zero crossings of the ZFF signal corresponds to instants of significant excitation i.e. Epochs.

Figure 3 shows the plot of speech signal, its ZFF signal and the epoch locations of the corresponding speech signal.



V. VERIFICATION

For verification of results obtained using ZFF algorithm the speech signal is passed through CMU Arctic database. This database is readily available on the internet and can be downloaded. This database extracts EGG signal from the input speech signal. EGG itself gives the instants of significant excitation i.e. epochs. Figure 4 shows that the epochs extracted using the ZFF method exactly matches with the EGG.

Figure 4 shows the speech signal, differenced EGG signal of the corresponding speech signal and epoch locations of speech signal obtained from the ZFF algorithm used in previous section.

VI. COMPARISON OF ZFF METHOD WITH LP BASED METHOD AND GD BASED METHOD

In this section the zero frequency filtering method is compared with the other two methods namely GD based method and LP based method. The LP based method for epoch extraction from speech signal uses block processing approach because of the use of this approach the accuracy of detection of epochs decreases. In the LP and GD based method uses LP residual signal which is computed using inverse filtering the speech signal i.e. by LP analysis. In ZFF method we do not need parameter selection but in LP analysis we need to select the parameters such as order of filter and the length of the window to compute the LP residual

All rights reserved by *www.ijirst.org*

signal. The ZFF method does not use periodicity property for epoch locations i.e. there is no need for the signal to be periodic but some methods use periodicity property of the signal for epoch extraction. The GD and LP based methods depends on the energy of the error signal and the energy of the speech signal. The GD based method does not works in case of noisy signal but ZFF method is robust against noise and gives accurate results.

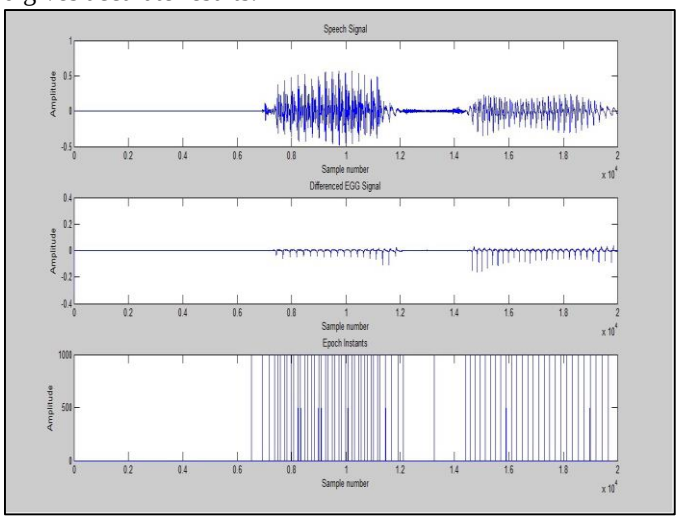


Fig. 4: Speech signal, Differenced EGG signal, Epoch locations of speech signal.

VII. CONCLUSION

The ZFF method is implemented for the epoch extraction from speech signal and is compared with LP based method and GD based method. It is observed that the ZFF method gives accurate epoch locations whereas LP based and GD based methods give approximate epochs because these methods use the block processing approach. ZFF method gives accurate epoch even under noisy conditions but GD based method works properly only in noise free environment. Since ZFF method gives accurate and reliable epoch locations it is useful for applications like pitch extraction from speech signal, in multispeaker separation, speech transformation and many more.

REFERENCES

[1] K. Sri Rama Murty and B. Yegnanarayan, "Epoch Extraction From Speech Signals", IEEE Trans. Audio Speech and Language Process, vol:16, no.8 ,Pages: 1602-1613, 2008.
[2] B. Yegnanarayana and Suryakanth. V .Gangashetty, "Epoch-based analysis of speech signals", Sadhana, vol. 36, Part 5, Pages:651–697, October 2011.
[3] Ma, Y., and Willems, L., "A Frobenius norm approach to glottal closure detection from the speech signal", IEEE Trans. Speech Audio Processing, vol. 2, Pages: 258-265, 1994.
[4] Ananthapadmanabha T V,B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval". IEEE Trans. Acoust. Speech Signal Process, vol:27,no.4,Pages: 309–319,1979 .
[5] Roel Smits and B. Yegnanarayana, "Determination of Instants of Significant Excitation in Speech Using Group Delay Function", IEEE Trans. Speech audio processing, vol 3,no.5, Pages: 325-333, 1995.
[6] K. Sreenivasa Rao, S. R. M. Prasanna, B.Yegnanarayana, "Determination of Instants of Significant Excitation in Speech Using Hilbert Envelope and Group Delay Function.", IEEE Signal Processing Letters, vol. 14, No. 10, Oct. 2007.
[7] K. Sreenivasa Rao and B. Yegnanarayana, "Prosody Modification Using Instants of Significant Excitation ", IEEE Trans. on Audio, Speech and Language Processing, vol. 14, No. 3, 2006.
[8] R. Kumara Swamy, K. Sri Rama Murty, and B. Yegnanarayana, "Determining Number of Speakers From Multispeaker Speech Signals Using Excitation Source Information " , IEEE Signal Processing Letters, vol. 14, No. 7, 2007.

All rights reserved by *www.ijirst.org*