

Improved Glottal Closure Instant Detector Based On Linear Prediction And Standard Pitch Concept

Cheol-Woo Jo¹, Ho-Gyun Bang¹, William A.Ainsworth²

¹ Dept. of Control and Instrumentation Engineering, Changwon National University, Korea
cwjo@sarim.changwon.ac.kr

² Dept. of Communication and Neuroscience, Keele University, U.K.
coa01@seq1.keele.ac.uk

ABSTRACT

This paper proposes an improved method of glottal closure instant detection using linear prediction and standard pitch concept. The main improvements are on its speed of computation and error reduction on position finding for the cases that were not possible or caused many errors using previous methods.

Our method can resolve the problems occurring in current methods to some extent. The false location detection rate is reduced to its inherent interpolation capability. Also the amount of computation is reduced. Another benefit from our method is that it does not need additional post processing to find peaks or smoothing of the pitch tracks. All is contained in itself. Also we compared results among three different kinds of linear prediction based pitch detectors.

1. INTRODUCTION

Detecting glottal closure instant(GCI) is one of the important problems from speech analysis including pitch synchronous analysis of speech and estimation of source characteristics speech from voiced segments.

Many different kinds of GCI detectors was suggested upto now. Many of the GCI detectors are based primarily on the linear predictive errors of speech signals. The main idea of such methods is that such error signal can indicate the most probable position of GCI. Also some improvements were made to emphasize the positions using a hilbert transform.

But there are some cases that error signal cannot indicate GCI position properly in such cases as high vowels. Also those methods require additional post processing to smooth the output tracks or to remove faulty detections.

In this paper we suggest a method which combines the linear prediction based GCI detectors and smoothing algorithms. As a result we can reduce the amounts of computation for the processing and can improve the performance of GCI detectors..

2. GCI DETECTION METHODS

There are many methods suggested about glottal closure instant detection. Most of them are based on the error signal of linear prediction. Other methods are based on maximum likelihood or laryngograph signal which is measured directly from the outside of human vocal fold.

Linear prediction based methods were EFLPR proposed by Yegnanaryana et. al. The method is based on the fact that prediction error signal is big at the location of glottal closures. This method works quite well. Generally it works for vowels or voiced consonants. But it sometimes can not show clear positions near probable GCI in such cases as high vowels or when the frequency of first formant and pitch frequency is very close or the same. So the hilbert transform was used to emphasize the position. And some post processing should be applied to adjust the delays of position.

Maximumlikelihood method was proposed by O'Shaughnessy et.al. This method is based on epoch filtering theory which is used for radar signal detection. This method also used the hilbert transformation to get emphasized position signal. And this method also shows some delays in positions.

Also methods using laryngograph are suggested but it is thought an inconvenient method due to attaching electrodes on the neck.

3. IMPROVED METHOD

In order to improve the performance of detection, we suggest additional processings to linear prediction based GCI detectors. Overall structure is shown on figure 1. The main differences with current method is in standard pitch scheme and change of S/V/UV decision stage.

3.1. Standard Pitch Scheme

We adopted standard pitch scheme to get the successive GCI locations from the speech signal. This scheme is based on the fact that the length of pitch period does not change drastically during the single segment of speech. In normal speaking conditions pitch periods do not double or halve during the limited length of the segment. So we obtained standard pitch candidate from the center

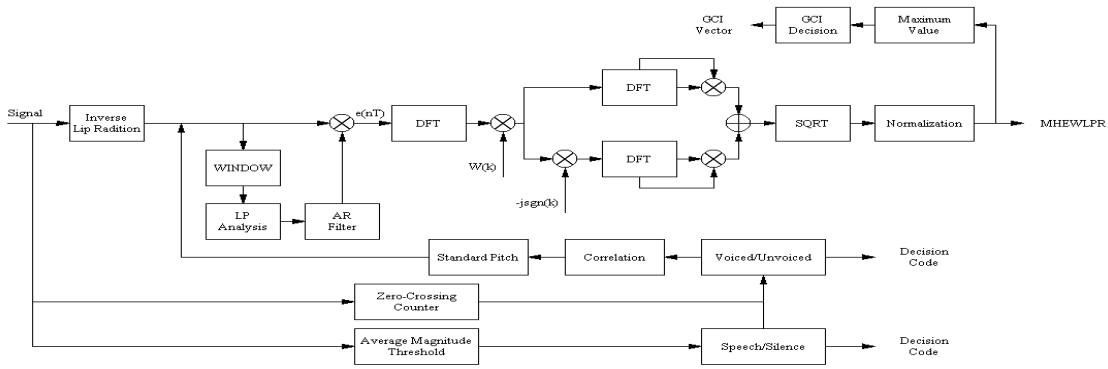


Figure 1: Schematic Diagram of Method

portion of the speech segment. Speech segment is decided from the scheme which will be mentioned at the next section.

By using this scheme we can get interpolation effect during the computation of GCI at the same time. Also we can reduce the amount of computations by reducing the size of the analysis frame size.

This scheme is described in figure 2.

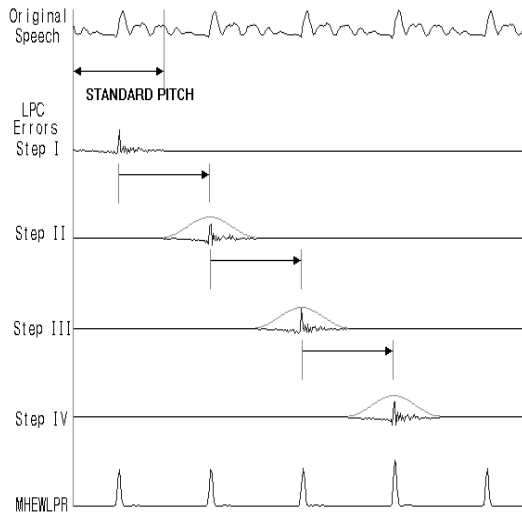


Figure2: Standard Pitch Concept

3.2. S/V/UV Decision

Silence, voiced and unvoiced decisions are made with simple parameters such as zero crossing rates and energy. Normally in conventional GCI detectors they use GCI detector output to make a S,V,UV decision. The reason why we made the

decision separately is that wrong decisions for GCI can lead to wrong decisions of S,V,UV. Main errors about S,V,UV decision occurs at the initial or final stages of sounds at which the energy levels are comparatively smaller than at the stationary parts of the speech signal. We think that error rate when deciding S,V,UV from the output of GCI detector will be similar to that when deciding from the energy or zero crossings of the speech signal. By deciding voiced segments of speech separately we can reduce the computational amounts.

3.3. GCI Detection Algorithm

We made decision from the MHEWLPR(Modified Hilbert Envelope from Windowed Linear Prediction Residual). From the voiced segments of the signal we get the linear prediction residual. Then from the center part of its spectrum the normalized hilbert envelope is obtained. On most stationary voiced segments we can get quite a good GCI pointers. But for some erroneous cases such as high vowels we often fails to find the location or just find false locations. The main reasons for that is the error signal is not enough to indicate GCI position in that signal. So our standard pitch scheme is proposed to help find such ambiguous positions. Applying the standard scheme is as follows. First, get the standard pitch value from the stationary center part of the voiced segments using autocorrelation function. In this case the frame length is chosen long enough to get an averaged pitch value. Using this value as a reference we start linear prediction analysis from the initial part of the voiced segment. The first GCI candidate is aquired based on conventional method. From the second GCI position we apply standard pitch scheme. By comparing previously obtained standard pitch with the new location candidate, we can get most probable location based on the fact that the human pitch period cannot change drastically due to the relatively slow movement of the physical organs. Also we apply a hanning window before analysis to get more emphasized GCI location. This makes finding GCI location easier.

Next, we remove any false locations from the forward and backward corrections. This scheme works as follows. First we start from the center part of the segment. By comparing

contiguous GCI positions and pitch intervals, we remove any provable halving or doubling between 80 and 120 % of standard pitch value. In figure 3 resulting corrected MHEWLPR and GCI indicator signal is shown with comparison to DEGG signal.

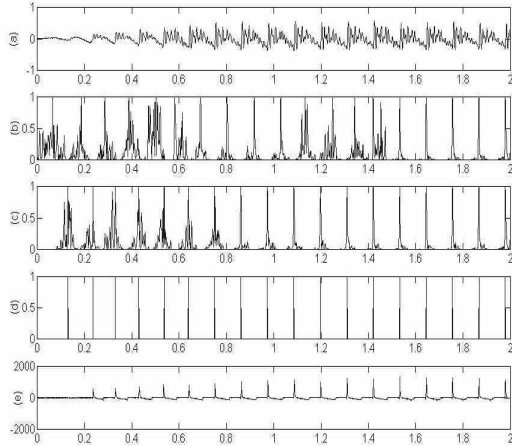


Figure 3: Analyzed Results from Initial /a/ sound

(a) original speech (b) MHEWLPR (c) Corrected MHEWLPR (d) GCI indicator (e) DEGG

4. EXPERIMENTAL RESULTS

Using the suggested algorithm we can obtain improved GCI locations from various speech materials. Then we compared the performance of the suggested method's result with the conventional EFLPR method's. We also compared the result with GCI locations from the differentiated laryngograph signal as a reference.

To prepare speech materials speech and laryngograph signal are sampled with 11 KHz sampling frequency. To sample speech we used laryngograph processor with DSP. The analysis is done with a SPARC-10 workstation using MATLAB.

List of speech samples are 10 pronounced korean numbers and Korean version of 'The North Wind and The Sun' and /Chang Won Dae Hak Gyo/ which means 'Changwon National University' in Korean.

Figure 4 shows comparisons of pitch trajectories from laryngograph output and suggested method. Pitch tracks follow closely to manually detected one. At the starting points and ending points it fails to find tracks. This is caused by false detection of speech segments from energy and zero crossings. But considering the fact that laryngograph itself often fails to find such positions, it is quite a natural phenomena.

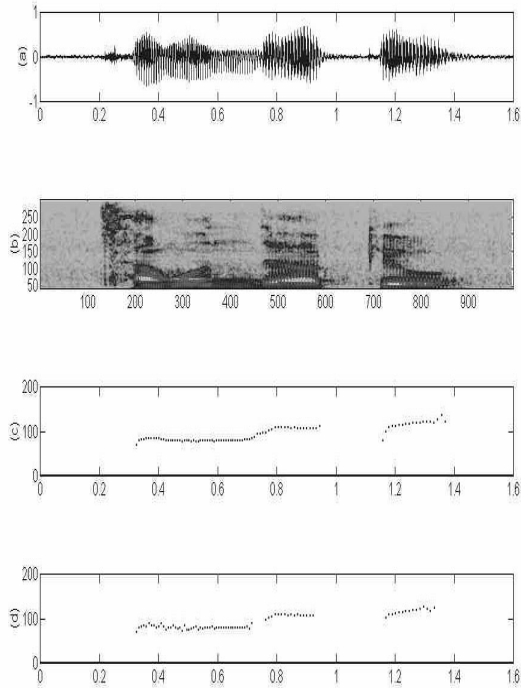


Figure 4: Speech /Chang Won Dae Hak Gyo/ and Pitch Tracks

(a) Original Speech (b) Spectrogram (c) Pitch Track from Laryngograph Signal (d) Pitch Track from Suggested Method

Figure 5 is comparisons among 6 kinds of signals. Original Speech, LPC residual, EFLPR, MLED, suggested method, DEGG are compared. From the figures we can see that our method indicates much explicit peak positions even at the initial transient part of the signal. The other method's output shows ambiguity at the transient parts.

Table 1. Shows average non-GCI detection error rate and average false GCI detection rates for both EFLPR and our method for 5 korean vowel signals. At table 1 we can see that our method gives better result and the result is much improved one from previous EFLPR method. From this result we can conclude that suggested successive method can help improve the performance of a conventional linear prediction based GCI detector.

	EFLPR	MHEWLPR
Non-GCI rate	16.02	3.4
False GCI rate	6.82	3.3

Table 1. Average Non-GCI detection and false GCI detection rates for two methods

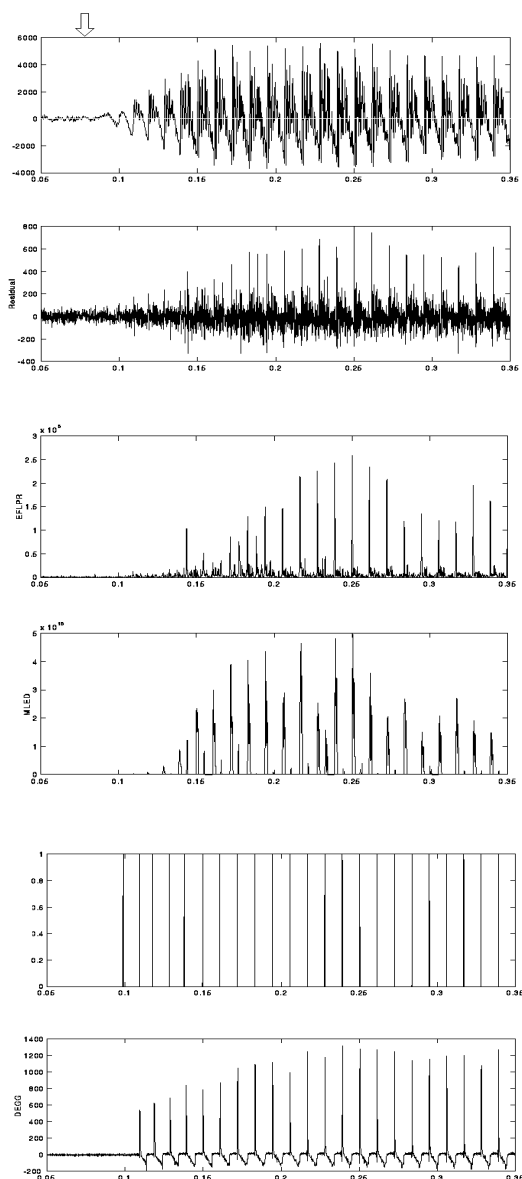


Figure 5. Comparisons of GCI signals

From above Original Speech , Residual Signal, EFLPR,MLED,MHEWLPR,DEGG

Most of the delays are concentrated on under 4 points delay.

Figure 6 shows number of delay points for test materials used.

5. CONCLUSIONS

In this paper we suggested an improvements to conventional linear prediction based GCI detectors. From the results we

can verify that the suggested method improves the performance of detecting GCI positions compared to the conventional linear prediction based method. This method is considered to be a good one for automatic GCI detections with implicit interpolation and correction capability.

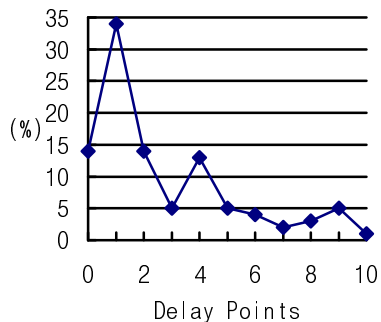


Figure 6. Delay points for test materials

6. REFERENCES

1. T.V.Anathapadmanabha, B.Yengnanaryana,"Epoch extraction from linear prediction residual for identification and close glottis interval",IEEE Trans. ASSP, Vol.27,No.4,pp.309-319,Aug.1979
2. Yang Ming Cheng, D.Oshaughness,"Automatic and Reliable Estimation of Glottal Closure Instant and Period",IEEE Tans ASSP,Vol.37,No.12,Dec. 1989
3. Ho-Gyun Bang, Cheol-Woo Jo,"A Study on the Detection of Glottal Closure Instant Using Sequential Linear Prediction",Proceedings of KSPC'94, Vol.7,1994
4. A.K.Krishnamurthy, D.G.Childers , " Two- Channel Speech Analysis ", IEEE Trans ASSP , Vol.34 , No.4, pp. 730 - 743, Aug.1986
5. D.Y.Wong,J.D.Markel,A.H.Gray,Jr.,"Least squares glottal inverse filtering from the acoustic speech waveform", IEEE Trans. ASSP, Vol.27, pp.350-355, Aug. 1979