Detection of Glottal Opening Instants using Hilbert Envelope

Ramesh K.[1], *S. R. M. Prasanna*[1] *and D. Govind*[2]

[1]Department of Electronics and Electrical Engineering,
Indian Institute of Technology Guwahati, Guwahati 781039, India,
[2] CEN, Amrita University, Coimbatore

{ kk.ramesh, prasanna } @iitg.ernet.in { d_govind } @cb.amrita.edu

10.21437/Interspeech.2013-10

Abstract

The objective of this work is to develop an automatic method for estimating glottal opening instants (GOIs) using Hilbert envelope (HE). The GOIs are secondary major excitations after glottal closure instants (GCIs) during the production of voiced speech. The HE is defined as the magnitude of complex time function (CTF) of a given signal. The unipolar property of HE is exploited for picking the second largest peak present in a given glottal cycle and hypothesize as glottal opening instant (GOI). The electroglottogram (EGG) / speech signal is first passed through the zero frequency filtering (ZFF) method to extract GCIs. With the help of detected GCIs, the secondary peaks present in the HE of dEGG / residual are hypothesized as GOIs. The hypothesized GOIs are compared with secondary peaks estimated from the dEGG / residual. The GOIs hypothesized by the proposed method show less variance compared to peak picking from dEGG / residual.

Index Terms: glottal opening instant, glottal closure instant, Hilbert envelope, difference EGG, difference speech, residual.

1. Introduction

The glottal closure instants (GCIs) and glottal opening instants (GOIs) are the primary and secondary major excitations, respectively in the glottal excitation signal during production of voiced speech [1]. Both theses instants are useful in the analysis and processing of voiced speech. For instance, they can be used as pitch markers for pitch synchronous analysis and prosody modification [2, 3, 4, 5]. Since GCIs are primary major excitations with associated large impulse strength, they are relatively easier to detect and several methods exist in the literature [6, 7, 8]. There are attempts to detect GOIs [9, 10, 11]. The most recent one proposes a method for detection of GOIs in the regions of interest (ROI) from linear predication (LP) residual [12]. The regions of interest for GOIs are identified by finding the mean smoothed signal from speech and considering negative zero crossing. After obtaining the ROI, the peak present in the LP residual is hypothesized as the GOI. Due to the bipolar nature of the residual, there will be ambiguities in the location of GOI and hence the variance associated with it. These issue can be addressed with the help of Hilbert envelope (HE) [13, 8] and hence the motivation for this work.

The HE of a given signal is defined as the magnitude of its complex time function (CTF) [14]. The given signal acts as the real part of CTF and Hilbert transform of given signal acts as imaginary part of CTF. The significance of HE for epoch extraction was demonstrated earlier [15, 8]. Since HE is a magnitude function and hence unipolar in nature. As a result, the ambiguities present in locating epochs from residual due to its bipolar nature can be minimized [15]. We may then expect prominent peaks around epochs in case of HE. This has been proposed earlier for locating GCIs in [15]. The same concept can now be extended, but for automatic detection of epochs, specifically GOIs.

The ground truth for the evaluation of methods for automatic detection of GCIs and GOIs from speech signal is needed. One approach is to consider the difference of electroglottogram (dEGG) and mark the impulses manually. This is labor intensive and error-prone method. Recently ZFF of electroglottogram (EGG) is proposed as an alternative for automatic detection of ground truth GCIs [7]. The present work uses the refined version of this method for obtaining ground truths for GOIs. Even though impulse-like nature is very prominent in the location of GCIs of dEGG, it is not so prominent in case of GOIs. The ZFF of dEGG provides GCIs, not GOIs. Therefore, a refined method is required for locating the GOIs automatically. The dEGG signal is processed initially by ZFF method for locating GCIs and are used as ground truths for further detection of GOIs. Between every pair of GCIs, the largest peak is identified as the GOI. The mean and variance values of GOIs provide the ground truth and can be used for evaluating the proposed method.

In the proposed method, the speech and EGG are passed through the ZFF for obtaining the reference GCIs. The speech signal is processed by linear prediction (LP) analysis to obtain LP residual. With the help of reference GCIs detected, methods are developed to locate secondary major excitation instants from dEGG and LP residual and hypothesized as GOIs. The secondary major excitations are also estimated using the proposed Hilbert envelope approach and hypothesized as GOIs. The detected GOIs using HE of LP residual are compared with that of LP residual. Based on this comparisons are drawn to demonstrate significance of HE for automatic detection of GOIs from speech signals.

The rest of the paper is organized as follows: Section 2 describes EGG signal, GCI, GOI, issues in detecting GOIs and HE method for detecting GOIs from dEGG. The proposed method for detection of GOI is described in Section 3. The experimental results and discussions are given in Section 4. The summary of current work and scope for future work are given in Section 5.

2. Estimationg of Glottal Opening Instants from EGG using HE

This section briefly describes the nature of EGG signal, GCI, GOI, HE method for the detection of GOI from dEGG.

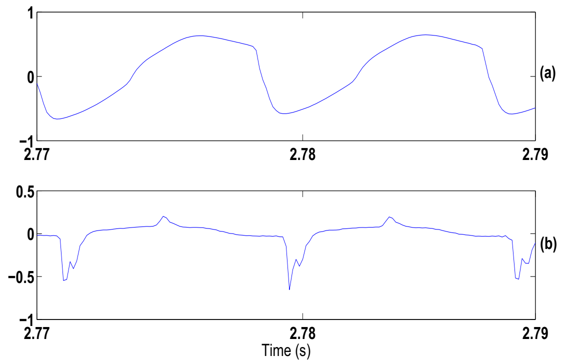

Figure 1: Glottal wave and its Derivative. (a) A voiced segment of EGG (b) its derivative indicating the discontinuities at GCIs and GOIs locations.

2.1. Nature of EGG Signal

The EGG signal gives information about contact area present between vocal folds during production of voiced speech. Figure 1(a) shows a segment of EGG signal. The large values in the EGG signal indicate opening phase of glottal cycle and hence less contact area between vocal folds. Similarly, the small values in the EGG signal indicate closing phase of glottal cycle and hence more contact area between vocal folds. Each of these phases in a glottal cycle are preceded by their onset and can be attributed to glottal opening instant and glottal closing instant respectively.

2.2. Glottal Closing and Opening Instants in EGG

The onsets of sharp impulse like discontinuities in the dEGG are known to be GCIs and the locations correspond to the peak with opposite polarity between successive GCIs are known as GOI [9]. Figure 1 plots a segment of EGG corresponds to a voiced speech and its derivative (dEGG). The negative peaks in the Figure 1 (b) indicate the GCIs and positive peaks having less amplitude compared to GCIs peaks, are indicated as the GOIs. The discontinuities associated with GOIs are not so prominent as that of the GCIs. Hence the automatic estimation of GOIs is a challenging task. In the present work the GCI locations are estimated initially and the positive peaks between successive GCIs are estimated to be as GOIs. As ZFF method provides accurate estimates of GCIs in speech and EGG, GCIs are estimated from the EGG using ZFF method.

In ZFF method, the dEGG is passed through a cascade of two zero frequency resonators [7]. As a zero frequency resonator (ZFR) is equivalent to a double integrator (cumulative sum), the zero frequency resonator output signal for dEGG is equivalent to four times integration or cumulative summation. This operation is represented mathematically by the input-output relation in Equation 1,

$$y(n) = -\sum_{k=1}^{4} a_k y(n-k) + x(n) \tag{1}$$

where a_1 = 4, a_2 = -6, a_3 = 4, a_4 = -1 and x(n) is the dEGG.

As the ZFR output is a polynomial growing function, the trend in the ZFR output is removed by local mean subtraction operation given in Equation 2,

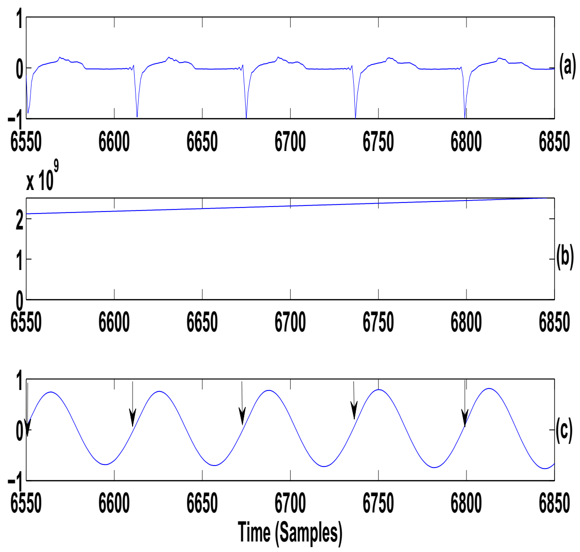

Figure 2: Estimation of GCIs from dEGG by ZFF. (a) A voiced segment of dEGG, (b) its ZFR output and (c) ZFFS of dEGG. The positive zero crossings of dEGG are marked using 'arrows'

$$\hat{y}(n) = y(n) - \bar{y}(n) \tag{2}$$

where $\bar{y}(n) = \frac{1}{2N+1}\sum_{n=-N}^{N} y(n)$ and $2N+1$ corresponds to average pitch period computed over a longer segment of dEGG. $\bar{y}(n)$ is termed as the zero frequency filtered signal (ZFFS). The negative to positive zero crossings in the ZFFS are hypothesized as GCIs. Figure 3 plots a segment of dEGG, ZFR output and ZFFS of dEGG. The estimated GCIs obtained by negative to positive zero crossings in the ZFFS corresponds to negative peaks in dEGG.

Once the GCIs are obtained by ZFF of dEGG signal, the GOIs are obtained by picking peaks between successive GCI locations. These GOIs estimated from dEGG peaks are used as reference GOIs for HE based GOI estimation.

2.3. Evidence of GCIs and GOIs in Hilbert Envelope of dEGG

Hilbert envelope is the magnitude of complex time function having signal and its Hilbert transformation as real and imaginary parts, respectively [14, 16]. The Hilbert envelope of a discrete time sequence $s(n)$ is given by,

$$h[n] = |s_a(n)|. \tag{3}$$

where $s_a(n)$ is complex time function can be computed as follows,

$$s_a(n) = s(n) + js_h(n) \tag{4}$$

where $s_h(n)$ is Hilbert transform of $s(n)$. The Hilbert transform is computed as

$$s_h(n) = IDFT(S_H(\omega)), \tag{5}$$

where

$$S_H(\omega) = \begin{cases} +jS(\omega), -\pi \le \omega < 0 \\ -jS(\omega), 0 \le \omega \le \pi \end{cases} \tag{6}$$

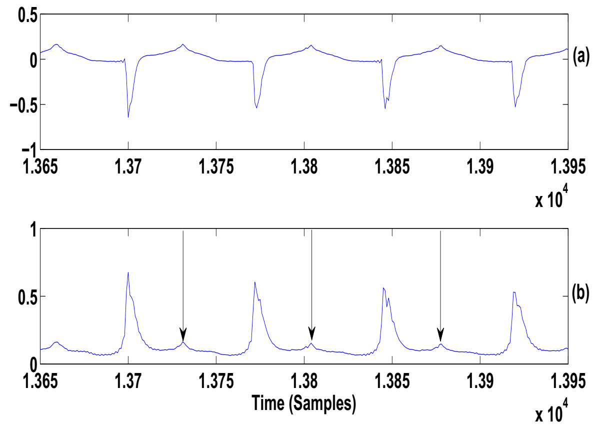

Figure 3: Evidences of GOIs in HE of dEGG. (a) A voiced segment of dEGG and (b) Corresponding Hilbert envelope. The secondary peaks in HE of dEGG indicate GOIs and are marked using 'arrows'

and $S(\omega)$ is DFT of $s(n)$. DFT refers to discrete Fourier transform and IDFT refers to inverse of DFT.

Therefore the magnitude of complex time function $s_a(n)$ is given by,

$$h(n) = \sqrt{s^2(n) + s_h^2(n)}. \quad (7)$$

Figure 3 plots a voiced segment of dEGG and corresponding HE of dEGG. The evidence of GCI and GOIs can be observed as the primary and secondary peaks in the HE and corresponds to the peaks in dEGG. Hence GOIs can be estimated from HE of dEGG by picking the secondary peaks. Since we have an efficient and reliable method like ZFF for estimating GCIs, the secondary peaks from HE of dEGG can be estimated by the knowledge of the estimated GCIs from dEGG. The GOIs are estimated to be locations corresponding to peaks between successive GCIs.

2.4. GOI estimation using HE of dEGG

The GOI performance is evaluated across CMU-Arctic database having simultaneous speech and EGG recordings [17]. The CMU Arctic database consists of three speakers (two males and one female) with 1132 phonetically balanced utterances for each speaker. All the utterances are down sampled to 8 kHz from the original sampling rate of 32 kHz. The GOIs estimated using HE of dEGG are compared with reference GOIs obtained by direct peak picking of secondary peaks between two successive GCIs in dEGG. The performance measures used for GOI estimation are GOI identification rate (IDR), GOI miss rate (MR), false alarm rate (FAR) and GOI identification accuracy. These performance measures are described in [18] for evaluating GCI estimation performance. The description of each measure is as follows:

- Larynx cycle: The range of sample (1/2) $(l_{r-1} + l_r) < n < (1/2)(l_{r+1} + l_r)$ where l_r, l_{r-1} and l_{r+1} are the current, preceding and succeeding reference GOI locations, respectively
- Identification Rate (IDR): The percentage of larynx cycles for which exactly one GOI is detected.

Table 1: Performance evaluation of GOI estimation by the ZFF of HE of dEGG.

CMU-Arctic Spkr	IDR (%)	MR (%)	FAR (%)	IDA (ms)
BDL	99.80	0.15	0.05	0.36
JMK	99.96	0.02	0.02	0.37
SLT	99.99	0.01	0.01	0.44
Tot. Avg	99.91	0.05	0.03	0.39

- Miss Rate (MR): The percentage of larynx cycles for which no GOI is detected.
- False Alarm Rate (FAR): The percentage of larynx cycles for which more than one GOI is detected.
- Identification Error (ζ): The timing error between reference and detected GOI in larynx cycles for which exactly one GOI was detected.
- Identification Accuracy (σ) (IDA): The standard deviation of identification error ζ. Small values of σ indicate high accuracy of identification.

Table 1 shows the GOI estimation performance for HE of dEGG. The effectiveness GOI estimation from HE of dEGG is confirmed by the higher IDR and IDA, and lower MR and FAR. The results presented in Table 1 confirms the significance of HE for the reliable estimation of GOIs. The merit of using HE is its unipolar nature and hence less ambiguity in locating the peaks. This is because, the direct peak picking from the signal may have some ambiguity due to multiple peaks present around the GOIs. Thus HE can be used for locating the GOIs.

3. Estimation of GOIs from Speech using HE

The GCI and GOI are not directly evident in the speech signal due to dominance of vocal tract information. For this, vocal tract and excitation source information needs to be separated. One way is to perform LP analysis and inverse filtering to obtain LP residual. For a proper LP order, the LP residual mostly contains source information. In particular, the large error regions during voiced speech can be attributed to GCIs and GOIs.

Figure 4 plots LP residual and HE of LP residual for a voiced segment of speech. The LP residual is obtained by 10^{th} order LP analysis of speech with frame size of 20 ms and a frame shift of 10 ms. The GCIs in LP residual are characterized by the largest impulse like discontinuities around GCI locations. The evidence of GOIs in the LP residual can be observed by weaker impulses between successive GCI locations. The evidence of GOIs in LP residual of Figure 4(b) can be observed by comparing the secondary peaks in dEGG plot shown in Figure 4(d). However, the peak locations are ambiguous due to the bipolar nature of LP residual. The approximate GOI locations can be estimated from the absolute of LP residual using the GCIs. Hence, the locations of largest impulse in the absolute of LP residual between successive GCIs are hypothesized as reference GOI locations.

Figure 4(c) plots HE of LP residual. The evidence of GOIs in HE can observed as secondary peaks between successive largest peaks which correspond to GCIs. By comparing Figure 4(b) and (c), it can be observed that secondary peak in HE of LP residual is more prominent and less ambiguous than secondary peaks of impulse location in LP residual. Also comparison with Figure 4(d) confirms that there is no much deviation of

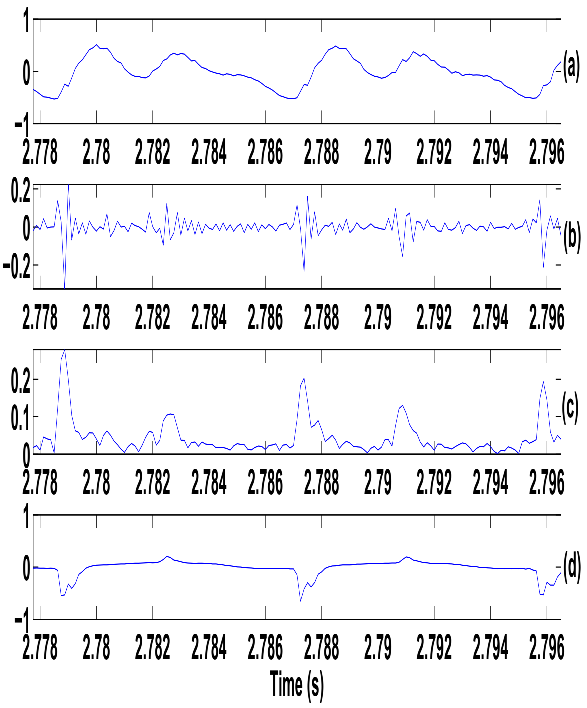

Figure 4: Evidences of GOIs in LP residual and HE of LP residual. (a) A voiced segment of speech, (b) LP residual, (c) the HE of LP residual and reference dEGG segment .

Table 2: Performance evaluation of GOI estimation by the ZFF of HE of LP residual.

CMU-Arctic Spkr	IDR (%)	MR (%)	FAR (%)	IDA (ms)
BDL	99.03	0.72	0.25	1.09
JMK	98.81	0.75	0.46	0.87
SLT	99.97	0.01	0.02	0.54
Tot. Avg	99.36	0.49	0.24	0.83

GOIs in HE of LP residual [15] with that of true GOI locations in dEGG.

Figure 5 plots the segment of LP residual, ZFFS obtained by ZFF of difference speech, estimated GCIs by the positive zero crossings of ZFFS, HE of LP residual for picking secondary peaks corresponds to GOIs and Estimated GOIs. The GCIs estimated by ZFF of speech are used for estimating GOIs from HE of LP residual. GOIs are estimated by picking peaks between successive GCI locations in the HE of LP residual. Figure 5(e) plots estimated GOI sequence.

Table 2, presents the IDR, MR, FAR and IDA obtained for the estimated GOIs when compared with reference GOIs obtained by picking the secondary impulse location between successive GCIs in the absolute of LP residual. It has to be observed that the IDR, MR, FAR and IDA are comparable with that obtained for the dEGG case.

The proposed method is compared with the existing methods in the literature reported in [12, 18]. The comparison results are given in the Table 3. The proposed method provides improved performance compared to these two recent methods. The proposed method is closely comparable with method in

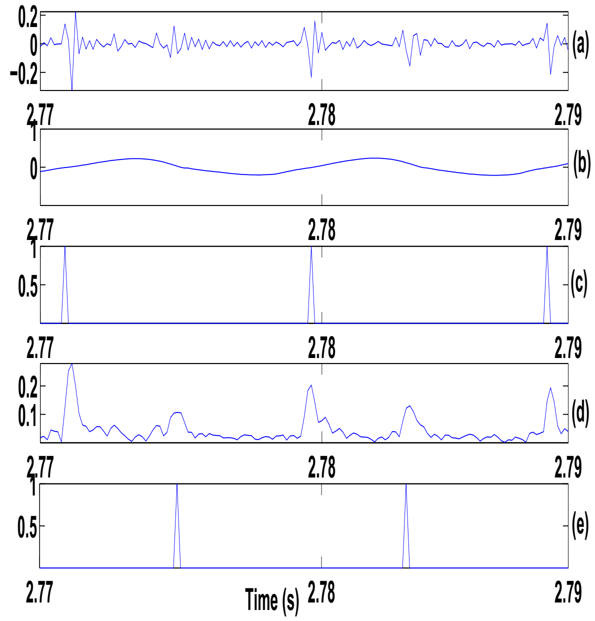

Figure 5: GOI estimation using HE of LP residual. (a) A voiced segment of LP residual, (b) the zero frequency filtered signal (c) Estimated GCIs from the positive zero crossings of ZFFS, (d) HE of LP residual (e) Estimated GOIs by picking secondary peaks from HE between successive GCIs .

Table 3: Comparative results of GOI estimation.

CMU-Arctic Spkr	Method	IDR (%)	MR (%)	FAR (%)	IDA (ms)
BDL	Proposed	99.03	0.72	0.25	1.09
	MBS+LPres	96.74	2.62	0.65	1.24
	DYPSA	96.30	1.68	2.01	1.12
JMK	Proposed	98.81	0.75	0.46	0.87
	MBS+LPres	98.50	0.99	0.50	1.05
	DYPSA	97.30	1.60	0.02	0.54
SLT	Proposed	99.97	0.01	0.01	0.54
	MBS+LPres	99.47	0.49	0.50	0.65
	DYPSA	98.26	0.34	0.86	0.65

[12], since this method also detects the GOIs from LP residual. The merit of the current method is avoiding locating region of interest and searching for peak. Also, the HE instead of LP residual reduces the ambiguities associated with the location as reported in [15]. Hence the improvement in the performance.

4. Summary and Conclusions

This work proposed a method for the detection of GOIs in EGG and speech using Hilbert envelope. The unipolar nature of HE is exploited for estimating the locations of GOIs. The proposed HE based method provides reduces ambiguity in the location of GOI and also provides improved performance.

The peak in the HE is better compared to LP residual peak, but still it may not actually correspond to the exact location of GOIs as indicated by the large value of IDA in case of speech. The future work should focus on reducing this IDA. Also methods may be developed for detection of GOIs from speech than its LP residual.

5. References

[1] T. F. Quatieri, *Discrete-Time Speech Signal Processing*. Pearson Education, 2004.

[2] T. F. Quatieri and R. J. McAulay, “Shape invariant time scale and pitch modification of speech,” *IEEE Trans. on Signal Process.*, vol. 40, no. 3, pp. 497–510, Mar 1992.

[3] E. Mourlines and J. Laroche, “Non-parametric techniques for pitch-scale and time-scale modification of speech,” *Speech Commun.*, vol. 16, pp. 175–205, 1995.

[4] K. S. Rao and B. Yegnanarayana, “Prosody modification using instants of significant excitation,” *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, pp. 972–980, May 2006.

[5] S. R. M. Prasanna, D. Govind, K. S. Rao, and B. Yenanarayana, “Fast prosody modification using instants of significant excitation,” in *Proc Speech Prosody*, May 2010.

[6] T. V. Ananthapadmanabha and B. Yegnanarayana, “Epoch extraction of voiced speech,” *IEEE Trans. Acoust., Speech and Signal Process.*, vol. 23, no. 6, pp. 562–570, Dec 1975.

[7] K. S. R. Murty and B. Yegnanarayana, “Epoch extraction from speech signals,” *IEEE Trans. Audio, Speech and Language Process.*, vol. 16, no. 8, pp. 1602–1614, Nov. 2008.

[8] K. S. Rao, S. R. M. Prasanna, and B. Yegnanarayana, “Determination of instants of significant excitation in speech using hilbert envelope and group delay function,” *IEEE Signal Processing Letters*, vol. 14, pp. 762–765, Oct. 2007.

[9] A. Bouzid and N. Ellouze, “Voice source parameter measurement based on multi-scale analysis of electroglottographic signal,” *Speech Communication*, vol. 51, pp. 782–792, 2009.

[10] N. Ellouze and A. Bouzid, “Open quotient measurements based onmultiscale product of speech signal wavelet transform,” *Research Lett. in Signal Processing, Hindawi*, vol. 7, pp. 1–5, 2007.

[11] N. Henrich, B. Doval, and C. dAlessandro, “Glottal open quotient estimation using linear prediction,” in *Proc. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, 1999.

[12] T. Drugman and T. Dutoit, “Glottal closure and opening instant from speech signals,” in *Proc. INTERSPEECH*, 2009.

[13] S. R. M. Prasanna and B. Yegnanarayana, “Extraction of pitch in adverse conditions,” in *Proc. ICASSP*, Montreal, Canada, May 2004.

[14] L. Cohen, *Time-Frequency Analysis: Theory and Applications*, S. P. Series, Ed. ser. Signal Processing Series. Englewood Cliffs: Prentice-Hall, 1995.

[15] T. V. Ananthapadmanabha and B. Yegnanarayana, “Epoch extraction from linear prediction residual for identification of closed glottis interval,” *IEEE Trans. Acoust., Speech and Signal Process.*, vol. ASSP-27, no. 4, pp. 309–319, 1979.

[16] D. Govind, S. R. M. Prasanna, and D. Pati, “Epoch extraction in high pass filtered speech using hilbert envelope,” in *Proc. Interpseech*, 2011.

[17] J. Kominek and A. Black, “CMU-Arctic speech databases,” in *in 5th ISCA Speech Synthesis Workshop*, Pittsburgh, PA, 2004, pp. 223–224.

[18] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, “Estimation of glottal closure instants in voiced speech using DYPSA algorithm,” *IEEE Trans. Audio, Speech and Lang. Process.*, vol. 15, no. 1, pp. 34–43, 2007.