Detection of Glottal Opening Instants using Hilbert Envelope

Ramesh K.\textsuperscript{1}, S. R. M. Prasanna\textsuperscript{1} and D. Govind\textsuperscript{2}

\textsuperscript{1}Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati 781039, India, \textsuperscript{2}CEN, Amrita University, Coimbatore

\{ kk.ramesh, prasanna \}@iitg.ernet.in \{ d.govind \}@cb.amrita.edu

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 hypothesizing 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 \cite{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 \cite{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 \cite{6, 7, 8}.

There are attempts to detect GOIs \cite{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 \cite{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) \cite{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) \cite{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 \cite{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 \cite{15}. We may then expect prominent peaks around epochs in case of HE. This has been proposed earlier for locating GCIs in \cite{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 \cite{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. Estimation 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.
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 GOIs [9]. Figure 1 plots a segment of EGG corresponds to a voiced segment of dEGG and GOIs locations.

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

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

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'.

\[ \hat{y}(n) = y(n) - \bar{y}(n) \]

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. \( \hat{y}(n) \) 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.

\[ h[n] = |s_n(n)| \]

where \( s_n(n) \) is complex time function can be computed as follows,

\[ s_n(n) = s(n) + js_h(n) \]

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

\[ s_h(n) = IDFT(S_H(\omega)) \]

where

\[ S_H(\omega) = \begin{cases} +jS(\omega), & -\pi \leq \omega < 0 \\ -jS(\omega), & 0 \leq \omega \leq \pi \end{cases} \]
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,

$$h(n) = \sqrt{s_{a}^{2}(n) + s_{x}^{2}(n)}.$$  \hfill (7)

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’

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

<table>
<thead>
<tr>
<th>CMU-Arctic Spkr</th>
<th>IDR (%)</th>
<th>MR (%)</th>
<th>FAR (%)</th>
<th>IDA (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDL</td>
<td>99.80</td>
<td>0.15</td>
<td>0.05</td>
<td>0.36</td>
</tr>
<tr>
<td>JMK</td>
<td>99.96</td>
<td>0.02</td>
<td>0.02</td>
<td>0.37</td>
</tr>
<tr>
<td>SLT</td>
<td>99.99</td>
<td>0.01</td>
<td>0.01</td>
<td>0.44</td>
</tr>
<tr>
<td>Tot. Avg</td>
<td>99.91</td>
<td>0.05</td>
<td>0.03</td>
<td>0.39</td>
</tr>
</tbody>
</table>

- 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 ($\zeta$): The timing error between reference and detected GOI in larynx cycles for which exactly one GOI was detected.
- Identification Accuracy ($\sigma$) (IDA): The standard deviation of identification error $\zeta$. Small values of $\sigma$ indicate high accuracy of identification.

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 10th 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 be 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

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,

$$h(n) = \sqrt{s_{a}^{2}(n) + s_{x}^{2}(n)}.$$  \hfill (7)

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’

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

<table>
<thead>
<tr>
<th>CMU-Arctic Spkr</th>
<th>IDR (%)</th>
<th>MR (%)</th>
<th>FAR (%)</th>
<th>IDA (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDL</td>
<td>99.80</td>
<td>0.15</td>
<td>0.05</td>
<td>0.36</td>
</tr>
<tr>
<td>JMK</td>
<td>99.96</td>
<td>0.02</td>
<td>0.02</td>
<td>0.37</td>
</tr>
<tr>
<td>SLT</td>
<td>99.99</td>
<td>0.01</td>
<td>0.01</td>
<td>0.44</td>
</tr>
<tr>
<td>Tot. Avg</td>
<td>99.91</td>
<td>0.05</td>
<td>0.03</td>
<td>0.39</td>
</tr>
</tbody>
</table>

- 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 ($\zeta$): The timing error between reference and detected GOI in larynx cycles for which exactly one GOI was detected.
- Identification Accuracy ($\sigma$) (IDA): The standard deviation of identification error $\zeta$. Small values of $\sigma$ indicate high accuracy of identification.

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 10th 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 be 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.

<table>
<thead>
<tr>
<th>CMU-Arctic Spkr</th>
<th>IDR (%)</th>
<th>MR (%)</th>
<th>FAR (%)</th>
<th>IDA (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDL</td>
<td>99.03</td>
<td>0.72</td>
<td>0.25</td>
<td>1.09</td>
</tr>
<tr>
<td>JMK</td>
<td>98.81</td>
<td>0.75</td>
<td>0.46</td>
<td>0.87</td>
</tr>
<tr>
<td>SLT</td>
<td>99.97</td>
<td>0.01</td>
<td>0.02</td>
<td>0.54</td>
</tr>
<tr>
<td>Tot. Avg</td>
<td>99.36</td>
<td>0.49</td>
<td>0.24</td>
<td>0.83</td>
</tr>
</tbody>
</table>

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.

Table, presents the IDR, MR, FAR and IDA 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 [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


