ROBUST DETECTION OF GLOTTAL ACTIVITY USING UNWRAPPED PHASE ELECTROGLOTTOGRAPHIC SIGNAL

Tanumay Mandal†, *K. Sreenivasa Rao*‡

†Advanced Technology Development Centre,
‡ Department of Computer Science & Engineering,
Indian Institute of Technology Kharagpur, India
tanumay.mandal10@iitkgp.ac.in, ksrao@iitkgp.ac.in

ABSTRACT

Glottal Activity is defined by the process of exciting the vocal tract system by the vibration of the vocal folds during speech processing. Detection of glottal activity refers to identifying the glottal instants present within a glottal cycle. GCI and GOI are the two important glottal instants of a glottal cycle. The DEGG signal provides significant peaks at those instants for normal voicing, but they are prone to error in detection for low strength EGG signal. In this paper, we mainly focus on the segments of EGG signal where the strength of the EGG signal is very poor and irregular in periodicity. The robust detection of the glottal instants in those segments of EGG signal will enhance the overall accuracy of the detection of glottal instants. In the proposed method, the unwrapped phase of the EGG signal has been used for the detection of the glottal instants. The phase of the signal has uniform characteristics throughout the signal which helps to detect the glottal instants robustly.

Index Terms— Electroglottographic signal, Glottal Instants, Unwrapped Phase.

1. INTRODUCTION

Electroglottography (EGG) is a well-known approach in view of analyzing the vocal-folds activity [1],[2]. In the proposed method, EGG signal is used to determine glottal activity by identifying glottal closing instant (GCI) and glottal opening instant (GOI). The time instant, when the vocal folds start attempt to close is known as GCI and the time instant, when the vocal folds start attempt to open or separate from each other is known as GOI. Identifying those instants can lead to better source modeling and parameterize the glottal activity that can be useful for various speech processing and bio-medical applications such as pitch and duration modification, speaking rate modification, pitch normalization, glottal synchronization, speaker modification and investigation of disorders in vocal folds. [3].

Over past decade, researchers have proposed various methods for identifying the glottal activity from EGG [3],[4], [5] and speech signal [6],[7],[8],[9],[10],[11]. Rothenberg et al., use a threshold of 50% crossing level of the EGG signal amplitude from a base line [12]. Howard uses the 3/7 of maximum EGG amplitude threshold for GOI detection [13]. The major drawbacks of above existing methods are the lack of accuracy for the GCI detection and a missing of GOI. Many approaches analyze the EGG signal by searching for spikes in DEGG signal [14],[15],[16] and compare their amplitudes with the threshold value to obtain an estimate of glottal activity during voiced speech.

The DEGG signals provide accurate detections of glottal instants during normal voiced speech but they are prone to error in different types of vulnerable cases voicing which include the region of transition in laryngeal mechanisms (LMs) and the end of voicing. Based on the length and thickness of the vocal folds and the muscular laryngeal tensions, human voice production mechanism is characterized by four types of distinctive LMs, namely M0 (vocal fry), M1 (modal), M2 (falsetto) and M3 (whistle) [17],[18]. The switching between LMs provides an observable transition at the signal level as well as at the auditory level [19]. In these transition regions, strength and shape of positive peaks of the DEGG signal are significantly disturbed which makes it difficult to unambiguously locate GCIs and GOIs [15].

Similarly, at the end of voicing, the strength of the voice signal becomes very low and sinusoidal shape is observed in the EGG signal. As a result, the DEGG signal provides weak and flat peak at the closing and opening phase of the glottal cycle. Hence, the detection of precise GCI and GOI location using the DEGG signal is erroneous during the end of voicing.

Hence, in this work, we propose phase based method for robust and accurate detection of glottal instants using EGG signal. The phase of EGG signal carries significant information which can be extracted from the phase data. Even though

The present work is carried out under the project entitled "Accurate analysis of vocal folds activity for speech and biomedical applications (AVS)" sponsored by Ministry of Human Resource Development (MHRD), Govt. of India.

the strength of EGG is very low and irregular during the cases mentioned above, the phase information will not be affected [20]. Hence, detection of instants using phase information of EGG signal will be highly robust and accurate. The measured phase is normally wrapped onto the range $-\pi$ to $+\pi$, which does not reflect the true relative phase value. Thus phase unwrapping is required before visualization or further processing.

2. PROPOSED METHOD

In this method, the phase of the EGG signal is explored for detection of glottal instants. The method contains several steps for detection of glottal instants. The steps for proposed method are as follows:

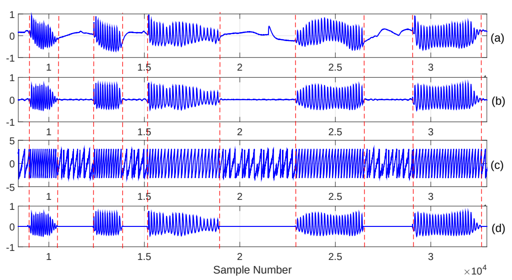

Fig. 1. Illustration of captured EGG signal and the pre-processed EGG signal. (a) Recorded EGG signal (b) Restored EGG signal after removing the low frequency components (c) Phase of the restored EGG signal (e) Pre-processed EGG signal

2.1. Pre-processing of the EGG signal

In EGG recordings, the real valued discrete time EGG signal can be expressed as the mixture of components of EGG waveform and the low frequency oscillation of the baseline produced by the slower movements of the other structures of the glottis. The pre-processing technique helps to extract the original EGG signal from low frequency noise like components and segregate the voiced region from the non-voiced regions of the captured EGG signal. Presence of low frequency components deliver uneven amplitude envelope in the EGG signal which influence the performance of the detection of glottal instants. High pass filter with cut-off frequency of 30 Hz has been used for removing the low frequency components from the captured EGG signal as the range of low frequency components of the signal lies between 0 Hz to 30 Hz. Figure 1(a) shows the recorded EGG signal. Figure 1(b) illustrates the EGG signal after removing the low frequency components from the recorded EGG signal. In next step, the phase of the EGG signal is computed using Hilbert transform to detect the non-voiced regions from the recorded EGG signal. Figure

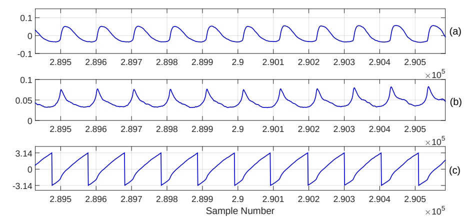

Fig. 2. Instantaneous amplitude envelope and phase of EGG signal computed over ten glottal cycles. (a) EGG signal, its (b) instantaneous amplitude envelope and (c) phase of the signal.

1(c) shows the phase signal of the EGG signal. The phase signal is uniformly wrapped from π to -π throughout the voiced regions but, it has non-uniform wrapping of the phase signal in case of non-voiced regions. This phenomenon helps to detect the non-voiced regions of the EGG signal. At the final stage of pre-processing, we replace those non-voiced regions from the EGG signal using zeros. Figure 1(d) illustrates the final stage of pre-processed EGG signal.

2.2. Unwrapped phase Extraction from EGG Signal

The analytic signal $e_a(n)$ corresponding to the EGG signal $e(n)$ is given by [21] $e_a(n) = e(n) + je_h(n)$ where $e_h(n)$ and $e(n)$ are the Hilbert pairs. The instantaneous amplitude envelope $a(n)$ and phase $\phi(n)$ of EGG signal can be computed from the analytic representation of EGG signal $e_a(n)$ using the following relations: $a(n) = \sqrt{e^2(n) + e_h^2(n)}$ and $\phi(n) = \tan^{-1}\left(\frac{e_h(n)}{e(n)}\right)$. Figure 2(a), 2(b) and 2(c) show the EGG signal, its corresponding amplitude envelope and phase of the EGG signal, respectively.

Here the phase $\phi(n)$ is wrapped as the phase response swings discontinuously by π radians when the frequency passes through a point where the phase crosses zero along the unit circle. The sudden transition of the phase (wrapped) distinguish the discontinuities present at the closing and opening instants of the vocal folds vibration. So, phase unwrapping is required to identify the discontinuities at closing and opening instants of the vocal folds vibration. The unwrapped phase function need not modify these discontinuities, but it is free to add or subtract any integer multiple of 2π in order to obtain the "best looking" discontinuity. The mathematical expression for phase unwrapping is defined by $\Delta\Phi(n) = \Phi(n) - \Phi(n-1)$. If $\Delta\Phi(n) \leq -2\pi$, then $\Phi_u(n) = \Delta\Phi(n) + 2\pi$, where $\Phi_u(n)$ is the unwrapped phase of the analytic signal. The reason for using phase unwrapping is mainly to perceive the significant discontinuities present in phase with respect to each glottal cycle. Figure 3(a) shows the pre-processed EGG signal for a normal voiced utterance. Figure 3(b) and 3(c) show the corresponding wrapped and unwrapped phase of the EGG signal. In wrapped phase, the phase signal is wrapped according to the period of glottal cy-

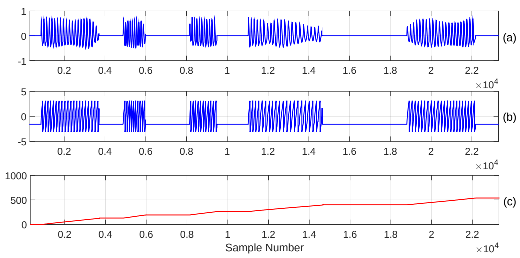

Fig. 3. Phase Extraction from the analytic signal (a) EGG Signal (b) Wrapped Phase (c) Unwrapped Phase

cle where, the unwrapped phase is monotonically increasing as the number of glottal cycles are increased.

2.3. GCI Detection using Unwrapped Phase

Fast and abrupt nature of the vocal folds during closing phase of EGG signal results in a sudden rise in the contact area of vocal folds which manifests a discontinuity in the phase of EGG signal. The discontinuity helps to detect the GCI instant of the EGG signal for each glottal cycle. In every EGG cycle, the unwrapped phase shows significant discontinuity present in the signal.

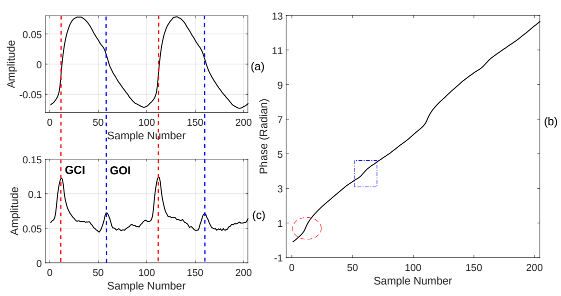

Fig. 4. Illustration of unwrapped phase and differentiated unwrapped phase (a) EGG signal (two cycles) (b) Unwrapped phase of the EGG signal (c) Differentiated unwrapped phase of the EGG signal

Figure 4(a) and 4(b) illustrate the EGG signal (two cycle) and corresponding unwrapped phase of the EGG signal. In figure 4(b), the red colored circle on the unwrapped phase is indicating the discontinuity present at the closing phase for every glottal cycle. This discontinuity on the unwrapped phase is due to sudden transition in glottal cycle at closing phase which basically indicates GCI of the glottal cycle. First order derivative of the unwrapped phase gives the significant positive high peaks on these discontinuities. Figure 4(c) illustrates the first order derivative of the unwrapped phase. In the figure 4(a) and 4(c), the vertical dash line (in red color) represents the GCI location for each glottal cycle. The peak of the first order derivative of the unwrapped phase signifies the GCI for each glottal cycle. The proposed method gives the accurate detection of the GCI by identifying the high positive peak for each glottal cycle.

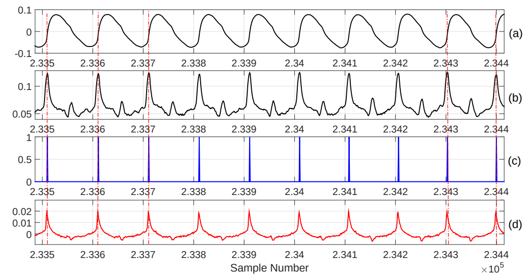

Fig. 5. Illustration of GCI detection using proposed method (a) EGG signal (b) Corresponding differentiated unwrapped phase (c) Proposed method detects the GCI location (d) DEGG signal

Figure 5 illustrates the detection procedure of GCI location using our proposed method. Figure 5(a) shows the preprocessed EGG signal. Figure 5(b) and 5(c) are indicating the corresponding first order derivative of the unwrapped phase of the EGG signal and GCIs locations are detected by using proposed method. Figure 5(d) shows the corresponding DEGG signal.

2.4. GOI detection using Unwrapped Phase

Glottal opening instant is the specific time instant at which the vocal folds initiate the activity of opening of their lower tissue layers from closed phase. This instant can be observed on the falling edge of EGG signal during each glottal cycle. As the rate of opening of the vocal folds is significantly low compared to rate of closing of the vocal folds, the change in DEGG signal at GOI (insignificant negative peak) is observed to be less significant, compared to change in DEGG signal at GCI (strong positive peak). Therefore, GOI detection from DEGG signal is not as accurate as GCI detection. In figure 4(b), the blue colored rectangle box on the unwrapped phase of the EGG signal is indicating the GOI of the glottal cycle. In figure 4(c), the secondary peak of the first order derivative of the unwrapped phase of the EGG cycle indicates the GOI location for each glottal cycle. The strength of the peak at GOI location in the figure 4(c) is suppressed due to high peak at GCI location. We clipped-off the peak present at the GCI location and emphasize the second peak of the first order derivative of the unwrapped phase of the EGG signal to identify the GOI instants. At first, we select the region where the closing phase of the glottal cycle is observed. In wrapped phase, the region lies between two consecutive zero crossings (first zero crossing from positive to negative transition of the wrapped phase and second zero crossing is negative to positive transition of the wrapped phase) of the phase for each glottal cycle. Figure 6 illustrates the process of removing the GCI region from the first order derivative of the unwrapped phase. Figure 6(a), 6(b) and 6(c) show the EGG signal and corresponding wrapped and first order derivative of unwrapped phase of that EGG signal. We mark the region which need to clip-off, with the help of two vertical red colored dash line "p" and "p'". We clipped-off the marked region ("p" to"p'") by replacing the value with zero which emphasizes the second peak of each glottal cycle. Figure 6(d) shows the first order derivative of the unwrapped phase of the EGG signal after GCI portion clipped-off. Now, we can visualize the precise peak presents at GOI location for each glottal cycle. The peak of the signal accurately signifies the GOI location for each glottal cycle in the EGG signal.

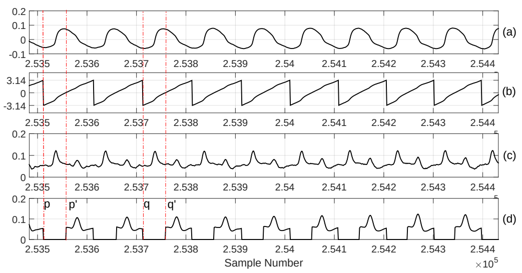

Fig. 6. Illustration of GOI detection using proposed method (a) EGG signal (b) Wrapped Phase of the EGG signal (c) First order derivative of unwrapped Phase of the EGG signal (d) First order derivative of unwrapped Phase of the EGG signal after GCI potion clipped-off

Figure 7 shows the detection procedure of GOI location of the glottal cycle using our proposed method. Figure 7(a) and (b) illustrate the EGG signal and corresponding first order derivative of unwrapped phase of the EGG signal after clip-off the closing phase. Figure 7(c) indicates the detection of GOI instant using the proposed method. In the figure 7(d) shows the DEGG signal.

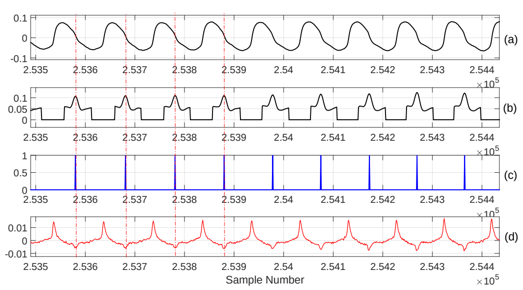

Fig. 7. Illustration of GOI detection using proposed method (a) EGG signal (b) First order derivative of unwrapped Phase of the EGG signal after GCi potion clipped-off (c) GOIs detection using proposed method (d) DEGG signal

3. ROBUSTNESS OF THE PROPOSED METHOD

At the end of voicing, proper discontinuities are absent at the closing and opening phase of the EGG signal. In DEGG signal, no precise peaks are located at those instants. The proposed method overcomes this limitation, by exploiting the presence of regularity of phase during the end of voicing. The phase of the EGG signal is robust in this type of voicing and provides precise glottal instants at the closing and opening phase of the glottal cycle. Figure 8 illustrates the glottal instants detection using proposed method for end of the voicing. Figure 8(a) illustrates the EGG signal. Figure 8(b) and 8(c) show the detection of GCIs and GOIs location at the end of voicing using the proposed method. The red dash line of figure 8(b) and 8(c) are indicating the GCIs and GOIs locations of the EGG signal. Figure 8(d) shows the corresponding DEGG signal. In DEGG signal, the strength of the peaks at those instants are very poor and almost flat at the opening instant which may provide an erroneous results for detections of glottal instants.

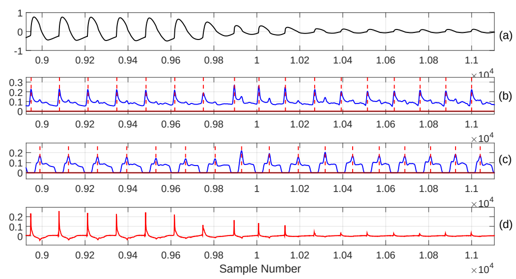

Fig. 8. Illustration of glottal instants detection in case of end of voicing (a) EGG signal (b) First order derivative of unwrapped Phase of the EGG signal(c) First order derivative of unwrapped Phase of the EGG signal after GCI portion clipped-off (d) DEGG signal

4. CONCLUSION

A novel approach has been proposed for detecting the glottal instants using EGG signal. The proposed method exploits the important characteristics of the phase of EGG signal, which is least influenced by strength or amplitude of EGG signal. Phase component of EGG signal mainly depends on the movement of glottis, and hence, we can accurately detect the glottal instants, even during the transition in LMs and end of voicing. Whereas, most of the state-of-the-art glottal instant detection methods may fail severely during above mentioned situations. The proposed method provides around 91% accuracy in detection of glottal instants for vulnerable cases of voicing whereas the existing methods provide on an average of 80-83% accuracy on these types of voicing. In case of vocal folds disorder patients', the proposed methods gives better performance than the existing methods. Based upon details, we believe that the proposed method has great potential in automated EGG signal analysis applications. In future direction, precise parameterization of EGG signal can be explored for synthesizing the vocal fold activity, which may be useful for tele-medicine applications.

5. REFERENCES

[1] D. G. Childers and J. Larar, “Electroglottography for laryngeal function assessment and speech analysis,” *IEEE Transactions on Biomedical Engineering*, vol. 31, no. 12, pp. 807–817, 1984.

[2] D. G. Childers, D. M. Hicks, G. P. Moore, L. Eskenazi, and A. L. Lalwani, “Electroglottography and vocal fold physiology,” *Journal of Speech, Language, and Hearing Research*, vol. 33, no. 2, pp. 245–254, 1990.

[3] B. Aicha and E. Noureddine, “Electroglottographic measures based on gci and goi detection using multiscale product,” *International Journal of Computers Communications and Control*, vol. 3, no. 1, pp. 21–32, 2008.

[4] P. B. Bachhav, H. A. Patil, and T. B. Patel, “A novel filtering based approach for epoch extraction,” *IEEE International Conference on Acoustics,Speech, and Signal Processing*, pp. 4784–4788, 2015.

[5] M. R. P. Thomas and P. Naylor, “The sigma algorithm: A glottal activity detector for electroglottographic signals,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 17, no. 8, pp. 1557–1566, 2009.

[6] K. S. R. Murty and B. Yegnanarayana, “Epoch extraction from speech signals,” *Journal of Speech, Language, and Hearing Research*, vol. 16, no. 8, pp. 245–254, 1990.

[7] 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, no. 10, pp. 762–765, 2007.

[8] T. Drugman, M. Thomas, J. Gudnason, P. Naylor, and T. Dutoit, “Detection of glottal closure instants from speech signals: A quantitative review,” *IEEE Transactions on Audio, Speech,and Language Processing*, vol. 20, no. 3, pp. 994–1006, 2012.

[9] A. I. Koutrouvelis, G. P. Kafentzis, N. D. Gaubitch, and R. Heusdens, “A fast method for high-resolution voiced/unvoiced detection and glottal closure/opening instant estimation of speech,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 2, pp. 316–328, 2016.

[10] S. R. Kadiri and B. Yegnanarayana, “Epoch extraction from emotional speech using single frequency filtering appoach,” *Speech Communications*, vol. 86, pp. 52–63, 2017.

[11] D. Thotappa and S. R. M. Prasanna, “Reference and automatic marking of glottal opening instants using egg signal,” *International Conference on Signal Processing and Communications (SPCOM)*, pp. 4260–4264, 2014.

[12] M. Rothenberg and J. J. Mahshie, “Monitoring vocal fold abduction through vocal fold contact area,” *Journal of Speech and Hearing Research*, vol. 31, pp. 338–351, 1988.

[13] D. M. Howard, “Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers,” *Journal of Voice*, vol. 9, no. 2, pp. 1212–1223, 1995.

[14] N. Henrich, C. dAlessandro, B. Doval, and M. Castellengo, “On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation,” *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1321–1332, 2004.

[15] N. Henrich, B. Roubeau, and M. Castellengo, “On the use of electroglottography for characterisation of the laryngeal mechanisms,” *Proceedings of the Stockholm Music Acoustics Conference*, pp. 6–9, 2003.

[16] N. Adiga and S. Prasanna, “Detection of glottal activity using different attributes of source information,” *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2107–2111, 2015.

[17] B. Roubeau, N. Henrich, and M. Castellengo, “Laryngeal vibratory mechanisms: The notion of vocal register revisited,” *Journal of Voice*, vol. 23, no. 4, pp. 425–438, 2009.

[18] A. Yamauchi, H. Yokonishi, H. Imagawa, K.-I. Sakakibara, T. Nito, N. Tayama, and T. Yamasoba, “Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging,” *Journal of Voice*, vol. 30, no. 2, pp. 205–214, 2015.

[19] S. R. Kadiri and B. Yegnanarayana, “Analysis of singing voice for epoch extraction using zero frequency filtering method,” *IEEE International Conference on Acoustics,Speech, and Signal Processing*, pp. 4260–4264, 2015.

[20] S. Kumar, T. Mandal, and K. S. Rao, “Robust glottal activity detection using the phase of an electroglottographic signal,” *Biomedical Signal Processing and Control*, vol. 36, pp. 27–38, 2017.

[21] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, “Discrete-time signal processing,” Prentice Hall,Upper Saddle River, NJ., 1999.