


Identification of glottal instants using electroglottographic signal for vulnerable cases of voicing

Tanumay Mandal , Krothapalli Sreenivasa Rao, Sanjay K. Gupta

Indian Institute of Technology Kharagpur, Kharagpur, India

 E-mail: tanum.dets@gmail.com

Published in Healthcare Technology Letters; Received on 26th September 2019; Revised on 17th April 2020; Accepted on 19th October 2020

Robust detection of glottal instants is essential for various speech and biomedical applications. Glottal closing and glottal opening are two crucial instants/epochs of a glottal cycle. The first-order derivative of the Electroglottographic (EGG) signal demonstrates important peaks at those locations for standard voicing, but the detection of glottal instants becomes erroneous when the peak to peak amplitude of the EGG signal is very low, irregular and unpredictable. In this work, a new efficient method is proposed for identification of glottal instants from the EGG signals including the segments of the signals where the signals are feeble with irregular periodicity. The overall accuracy of detection will be enhanced by identifying the glottal instants for the whole part of the signal including the vulnerable segments of signal. As the phase of a signal is uniform in nature, the phase information of the EGG signal has been explored to detect glottal instants accurately. Under low strength of the EGG signal, the proposed method remarkably has better performance compared to the existing instants detection methods and for pathological EGG signal, the detection accuracy of glottal instants is better than other existing methods.

1. Introduction: Analysis of the activities of the vocal folds is essential for various speech and biomedical applications like speech reverberation, speech synthesis, speaker recognition, identification of vocal folds disorders and so on [1–6]. Electroglottography (EGG) is a well known non-invasive approach for analysing the activity of the vocal folds [7, 8]. In this process, a high frequency, about $f \simeq 1$ MHz, modulated current of low voltage and amperage passes between two electrodes placed on the surface of throat at the level of alae of the thyroid cartilage. The vibrating behaviour of the vocal folds at the time of producing speech is well captured by EGG signal. Vibratory movements of vocal folds are analysed by measuring the variation of electrical admittance into the electrodes. The value of admittance becomes maximum, when the vocal folds are fully closed. Similarly, the value of admittance is minimum, when the vocal folds are at open position. In this method, EGG signals are used to analyse the glottal activity by spotting two very important glottal instants: glottal closing instant (GCI) and glottal opening instant (GOI) within an EGG/glottal cycle. At the time when two vocal folds initiate to come close together from their abduction position is called as GCI instant, similarly, at the time when both vocal folds are separating from each other from adduction position is called as GOI instant. GCI is mainly used in speech processing applications, like voiced non-voiced classification, source parameter estimation and speech synchronisation. GOI is helpful for voice quality estimation, voice and speaker characterisation [1]. In first order derivative of EGG (DEGG) signal, the positive high peak is indicating the GCI location and negative low peak is indicating the GOI location over an EGG cycle.

In state of the art, the researchers had proposed many methods and algorithms to determine the instants from EGG [1, 2, 9] and speech/acoustic signal [3–5, 10–13]. The main disadvantages of the existing methods were less accuracy in detection of GCIs and false detection of GOIs for the vulnerable cases of voicing. In state of the art, many methods detected the glottal instants by observing the spikes in the DEGG signal [14–16].

In [13], the authors proposed a method for the detection of GCI in degraded speech based on single-frequency filtering. In this method, they produced a noise compensate envelope. After that, they normalised the envelope and extracted the variance contour. The lowest slope value of the variance contour determined

the GCI. The performance of the proposed method improved significantly with the incorporation of noise compensation. In [17], the centre of gravity (CoG)-based method is used for detection of glottal instants. The CoG-based glottal instants detection method consists of four steps: linear prediction residual (LPR) extraction, Hilbert envelope computation, CoG computation and negative zero crossing point determination. The method assumes that the large peaks in the LPR signal correspond to the instants of significant excitation in voiced part of the signal. From results, it can be concluded that the CoG method detects all the instants in the voiced part meanwhile, it produces a large number of false detections for non-voiced part of the signal.

The method proposed in [6, 9] is based on the variational mode decomposition (VMD) and autocorrelation features for automatic identification of the glottal instants using EGG signal. In this method, first, a candidate EGG feature signal is generated based on centre mode frequency of the EGG signal. In the next stage, the glottal instants are identified using the zero crossing of the candidate EGG feature signal. Autocorrelation technique is used as a post processing technique to reject non-glottal instants from non-voiced part of the signal. In [18], SIGMA algorithm is proposed for proper detection of glottal instants. The authors used multi-scale product and group delay function to identify the glottal instants within a glottal cycle. SIGMA method demonstrates very impressive performance in normal voicing, though it needs to change some parameters to provide better performance. In many of the existing studies, the authors used DEGG signal as a reference for GCIs and GOIs detection. GCI is defined by the high positive peak of the DEGG signal and GOI is defined by the comparatively low flat negative peak of DEGG signal [14]. During normal utterance, the peaks of the DEGG signal indicate the actual position of the GCI and GOI instants. However, at the time of vulnerable cases of voicing, the peaks of the DEGG signal provide ambiguous indication of glottal instants.

Here, the vulnerable cases of voicing mainly refer to the regions/segments of voice transition in laryngeal mechanisms (LMs), and the low voice or the end of voicing. Depending on the thickness and length of the folds and the muscle tensions of laryngeal parts, the production of the human voice can be characterised into four groups of distinctive LMs like: M0 (vocal fry voice), M1 (modal voice), M2 (falsetto voice) and M3 (whistle voice) [19, 20].

The transition at the acoustics and EGG signal as well as the auditory level can be observed during the shifting between the LMs [15]. During speech production, M1 and M2 are the mostly used LMs. Hence, in the present work, switching between LMs has been studied and referred as the transition in LMs. In the transition regions, the shape and strength of the peaks of the DEGG signal are significantly disturbed [15]. Furthermore, the transition in LMs region, two closely spaced peaks (either positive or negative) in DEGG signal are observed. Due to ‘zipper like’ closing (or opening) of the vocal folds in the direction of the anterior–posterior dimension, the peaks are located very closely with equal amplitude. Generally, the closing of vocal folds begin at the anterior position and its move up to the posterior position to make a complete closure of vocal folds with a small amount of time delay. This type of behaviour of the vocal folds creates multiple peaks (mainly two) in DEGG signal. The presence of multiple closely spaced peaks in the DEGG signal during closing and opening phase makes it difficult to unambiguously locate glottal instants. At end of voice in a sentence or speech, the strength/amplitude of the EGG signal becomes very insignificant and low-frequency sinusoidal cycles are observed instead of normal EGG cycles. The discontinuities at the glottal instants are missing for this type of voice. Hence, the detection of the glottal instants is ambiguous for end of voice section or low voicing condition.

In this work, the phase information of the signal is examined for accurate identification of glottal instants (GCIs and GOIs). The phase signal carries significant information for identification of glottal instants. The phase information of the signal remains unaltered even though the strength of the signal is irregular and weak [21]. So, the detection of the instants using the phase information is highly robust and applicable. Normally, the extracted phase signal is wrapped onto the range of $-\pi$ to $+\pi$, which basically suppresses the discontinuity present in the signal. In the proposed work, the phase signal is unwrapped for visualisation of the actual discontinuities and for the further processing. The organisation of the paper is as follows: analysis of the proposed method and detection of glottal instants are discussed in Section 2. Evaluation process and results are discussed in Section 3. Summary of the paper is discussed in Section 4.

2. Proposed method: In this paper, the phase information of the EGG signal has been analysed for detection of glottal instants in automatic way. The explanations of the method have been discussed in [22]. In [22], we have discussed the method for identifying the glottal instants using the unwrapped phase of the EGG signal in detail. In this paper, the detail evaluation of the proposed method using the two modal databases (CMU-Arctic database and Apla wd database) along with vocal folds disorders patients’ database has been analysed throughly. The steps of the proposed method are discussed briefly in the following sections.

2.1. Pre-processing of the signal: At the time of EGG signal recordings, a low-frequency signal is coupled with the original EGG signal due to slower movement of the other structures of the glottis. To decouple the low-frequency elements from the EGG signal, and to get the actual noise-free EGG signal, a pre-processing technique is applied into the captured EGG signal. In this technique, all the silence and non-voiced segments are suppressed in the captured EGG signal. In the first step of pre-processing, the low-frequency elements are eliminated from the recorded EGG signal, and then non-voiced segments of the signal are marked and suppressed into zero for the captured EGG signal. A high pass filter is used to remove the low-frequency signal from the captured EGG signal. The cut-off frequency of the filter is kept as 30 Hz. Fig. 1 demonstrates the techniques of pre-processing the EGG signal. Fig. 1a shows the recorded EGG signal. Fig. 1b illustrates the EGG signal after removing the low-frequency components from the recorded EGG

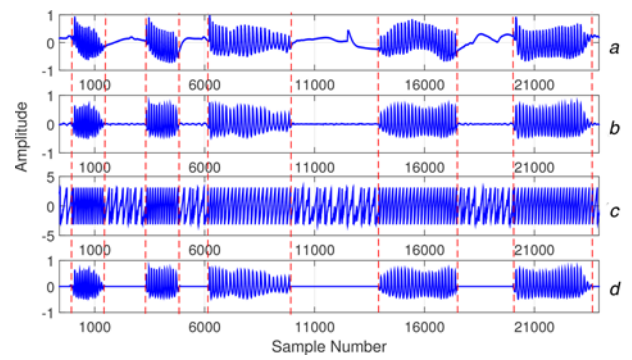


Fig. 1 Illustration of recorded EGG signal and the pre-processed EGG signal

- a Recorded EGG signal
- b Restored EGG signal after removing the low-frequency components
- c Phase of the restored EGG signal
- d Pre-processed EGG signal

signal. In the second step, we detect the non-voiced segments of the EGG signal and replace those segments from the signal by introducing zeros on those segments to enhance the performance of the proposed method. The phase part of the EGG signal helps to segregate the non-voiced segments from the recorded EGG signal. The phase of the signal is computed using the Hilbert transform. Fig. 1c shows the phase signal of the EGG signal. The phase of the EGG signal has uniform periodic nature for the voiced segments of the signal, whereas it shows irregular and noise like behaviour for the non-voiced segments of the EGG signal. The phase of the signal is uniformly wrapped from π to $-\pi$ throughout the voiced segments, but, it has non-uniform wrapping of the phase signal in case of non-voiced segments. This phenomenon helps to detect the non-voiced segments of the EGG signal. The period of wrapping of the phase signal is uniform and steady throughout the voiced segments whereas the property fails for the non-voiced segments. The non-voiced segments of the EGG signal are identified from the EGG signal by measuring the periodicity of wrapping the phase signal. After detection of the non-voiced segments, the segments are multiplied by zero to suppress the segments from the signal. Fig. 1d illustrates pre-processed EGG signal.

2.2. Extraction of unwrapped phase from the EGG signal: In this method, the Hilbert transform is used to make the EGG signal ‘analytic’ for extraction of instantaneous amplitude envelope and phase information of the EGG signal. Hilbert transform allows an asymptotically exact reconstruction of non-stationary signals. The analytic signal of the recorded EGG signal $e(n)$ is given by [23]

$$e_a(n) = e(n) + je_h(n) \quad (1)$$

here $e_h(n)$ and $e(n)$ are the Hilbert pairs of the EGG signal. $e_h(n)$ is the Hilbert transform of $e(n)$. Here, $e(n)$ is a real EGG signal. For extracting the phase information from the EGG signal $e(n)$, we performed the Hilbert transform on this signal to make the signal in complex form. $e_a(n)$ is the complex (analytic) form of the real EGG signal $e(n)$. The expression of the Hilbert transform is

$$e_h(n) = HT[e(n)] \quad (2)$$

The Hilbert transform $HT[e(n)]$ of a signal $e(n)$ is defined as

$$e_h(n) = e(n) * \frac{1}{\pi n} \quad (3)$$

The Hilbert transform of $e(n)$ is the convolution of $e(n)$ with the signal $(1/\pi n)$. The integral form of the transformation is

$$e_h(n) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e(\tau)}{n - \tau} d\tau \quad (4)$$

Hilbert transform is properly defined as the Cauchy principal value of the integral in (4), whenever this value exists. The Cauchy principal value is defined for the integral in (4) as

$$e_h(n) = \frac{1}{\pi} \lim_{\epsilon \rightarrow 0} \left(\int_{n-(1/\epsilon)}^{n-\epsilon} \frac{e(\tau)}{n - \tau} d\tau + \int_{n+\epsilon}^{n+(1/\epsilon)} \frac{e(\tau)}{n - \tau} d\tau \right) \quad (5)$$

the Cauchy principal value is obtained by considering a finite range of integration that is symmetric about the point of singularity, which excludes a symmetric subinterval, taking the limit of the integral as the length of the interval approaches ∞ while the length of the excluded interval approaches zero. Now, the analytic form of the EGG signal $e_a(n)$ can compute the amplitude envelope $a(n)$ and the phase $\phi(n)$ of the EGG signal using the following relations

$$a(n) = \sqrt{e^2(n) + e_h^2(n)} \quad (6)$$

$$\Phi(n) = \tan^{-1} \left(\frac{e_h(n)}{e(n)} \right) \quad (7)$$

here, $\Phi(n)$ is a wrapped phase. The phase is wrapped by 2π radians for every EGG cycle at the time when the phase cuts the zero line in the unit circle. The derivative of the wrapped phase signal overpower the discontinuities present at the glottal instants due to sharp transition of the phase for every cycle of the EGG signal. To overcome this problem and to identify the discontinuities present at the location of glottal instants, the phase signal needs to unwrap as the unwrapped phase does not contain the sharp transition of 2π radians for every EGG cycle. The unwrapped phase signal is free to add or subtract any integer multiple of 2π to make continuous for each cycle of the EGG signal and to obtain the 'best looking' discontinuities. The mathematical expressions for unwrapping the phase are defined by

$$\Delta\Phi(n) = \Phi(n) - \Phi(n - 1) \quad (8)$$

If $\Delta\Phi(n) \leq -2\pi$, then

$$\Phi_u(n) = \Delta\Phi(n) + 2\pi \quad (9)$$

where $\Phi_u(n)$ is defined as the unwrapped phase of the EGG signal. The actual reason for choosing the unwrapped phase is to maintain the certain discontinuities which are present at the glottal instants (GCI and GOI) for each EGG cycle. In the unwrapped phase, phase of the EGG signal is monotonically increasing as the number of the glottal cycles are increased. Fig. 2 illustrates the EGG signal with its wrapped and unwrapped phase of the EGG signal. Fig. 2a shows the EGG signal (ten cycles). Fig. 2b shows the corresponding wrapped phase and Fig. 2c shows the unwrapped phase of the EGG signal.

2.3. Detection of glottal instants: The abrupt change in the contact area of the vocal folds at the closing phase provides a discontinuity in the unwrapped phase of EGG signal. Similarly, during the opening phase, a minor discontinuity present in each glottal cycle. The discontinuities signify the glottal instants (GCI and GOI) present in each cycle of the EGG signal. However, the unwrapped phase of the EGG signal displays the discontinuities which are located at each glottal instant of the EGG cycles. Fig. 3a illustrates the EGG signal (two cycles). Fig. 3b shows the unwrapped phase of the EGG signal. Here, the phase signal is monotonically increased from 0 to 4π (2π for each cycle) for

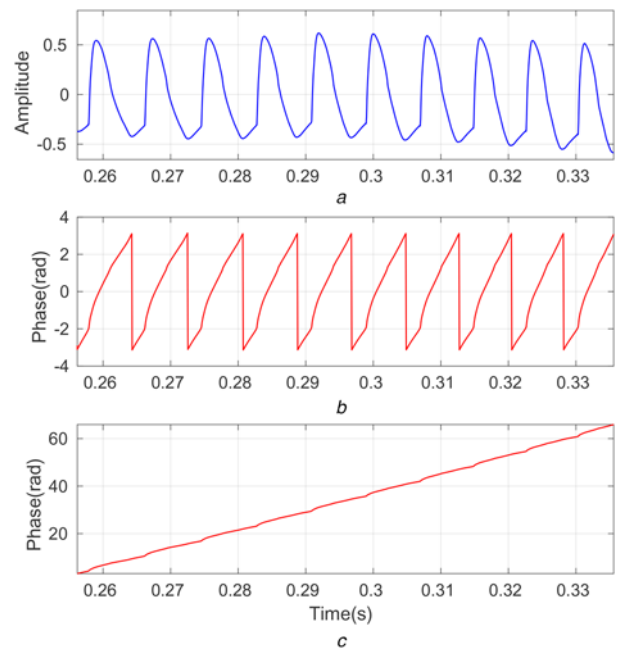


Fig. 2 Illustration of
a EGG signal (ten cycles), corresponding
b Wrapped phase
c Unwrapped phase

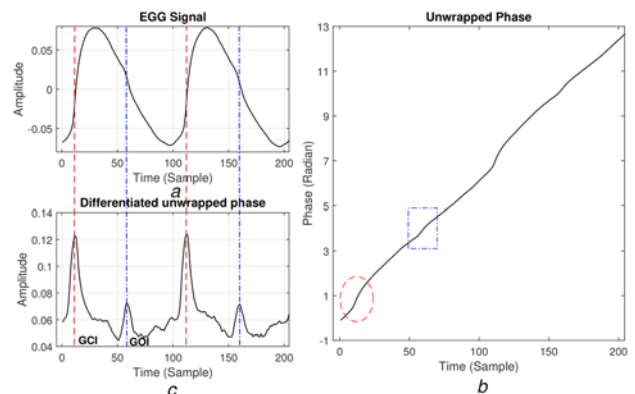


Fig. 3 Visualisation of identification of glottal instants using proposed method

- a Captured the EGG signal (two cycles)
- b Corresponding unwrapped phase
- c Corresponding derivative of the unwrapped phase of the EGG signal

two cycles EGG signal. For the first cycle (0–100 sample) of the EGG signal, two discontinuities (marked with red circle and blue rectangular box in Fig. 3b) are observed on the unwrapped phase signal. The two discontinuities represent the closing and opening instants of the each cycle of the EGG signal. Between the discontinuities, the first discontinuity has major transition compared to the second discontinuity for each cycle of the EGG signal. In the unwrapped phase of EGG signal, the first discontinuity (marked with red dash circle) signifies the GCI, whereas the second discontinuity (marked with blue dash rectangular box) signifies the GOI for each EGG cycle. Fig. 3c illustrates the first order derivative of the unwrapped phase of the EGG (DUPEGG) signal. In this Figure, the peaks for each cycle of the EGG signal are observed according to the presence of the discontinuities in the unwrapped phase of the signal. In Fig. 3c, the signal has two positive peaks with different strength for each glottal cycle. Between these two peaks, the

strength of the first peak is more compared to the second peak of the signal for each glottal cycle. In each cycle of the DUPEGG signal, the first high positive peak indicates the GCI location of the EGG signal. By identifying the highest peak of each cycle of the signal demonstrated in Fig. 3c, we can determine the GCI of the EGG signal accurately. The vertical red coloured dash line at the high positive peak of the DUPEGG signal in Fig. 3c perfectly coincides with the GCI location of the EGG signal in Fig. 3a, which confirms the accurate detection of GCI of the EGG signal using the proposed method.

The transition of vocal folds at the opening phase is smoother compared to the closing phase as a result, opening instant exhibits a minor discontinuity in the EGG signal as well as in the DEGG signal. As a result, detection of GOIs from the DEGG signal is ambiguous. The rectangular box (blue colour) on the unwrapped phase of the signal in Fig. 3b is indicating very small discontinuity of the unwrapped phase of the EGG signal. The discontinuity of the unwrapped phase of the signal is due to transition of the vocal folds from adduction position to start of opening/separating (this is known as GOI of EGG cycle) of the vocal folds. Comparatively lower strength positive peak of each cycle of DUPEGG signal as demonstrated in Fig. 3c, signifies the GOI of the EGG signal. We need to emphasise the secondary low strength peak of the signal for precise detection of the GOI of the EGG signal. For this purpose, the region of the GCI location is suppressed/clipped-off from the derivative of the unwrapped phase of the EGG signal. Following techniques are used to detect the GOI accurately within a glottal cycle using the proposed method. At first, the GCI location is identified for each cycle of the EGG signal. After that, we select the region of one millisecond (16 samples as the sampling frequency is 16 KHz) before and after the GCI location in the DUPEGG signal. In Figs. 4c and d, the vertical black coloured dash lines for each EGG

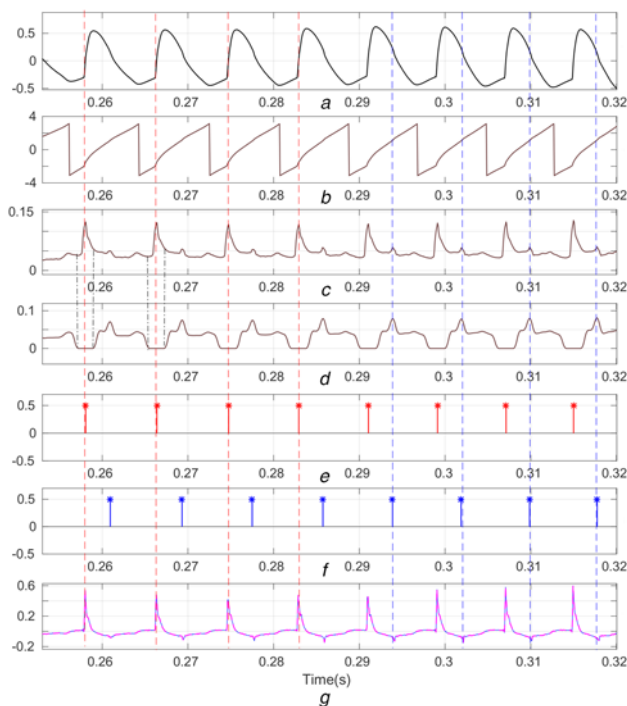


Fig. 4 Illustration of detection of glottal instants of the EGG signal using proposed method

- a Raw EGG signal (eight cycles)
- b Corresponding wrapped phase
- c Derivative of wrapped phase
- d Derivative of unwrapped phase signal after clipped-off the closing regions
- e Detected GCI instants
- f Detected GOI instants
- g Corresponding DEGG signal

- **Input:** EGG signal $e(n)$.
- **Output:** Glottal Instants: GCIs, GOIs

- 1 Preprocess the EGG signal by removing the anomalies like low frequency coupled signal, non voiced part of the signal etc.
- 2 Use Hilbert transform to get the phase of the EGG signal.
- 3 Extract the unwrapped phase and take the first order derivative of the unwrapped phase (DUPEGG) signal.
- 4 The primary peak (higher one) of the DUPEGG signal for each glottal cycle determines the GCI of the glottal cycle.
- 5 Detect the closing phase of each glottal cycle and remove the portion from the DUPEGG signal to emphasize the secondary peak of the signal for each glottal cycle.
- 6 The emphasized secondary peak of the DUPEGG signal determines the GOI of the glottal cycle.

Fig. 5 Algorithm: steps of the proposed phase-based method for detecting the significant glottal instants (GCIs and GOIs)

cycle indicate the selected regions. In next step, the selected regions are suppressed to zero to emphasise the secondary peak of the DUPEGG cycle. Now, in Fig. 4, the vertical blue coloured dash lines indicate the GOI locations of the EGG signal. Fig. 4a illustrates the EGG signal and Fig. 4b illustrates the phase of the corresponding EGG signal. Fig. 4c indicates the first order DUPEGG signal. The Figure shows the two positive peaks for each EGG cycle. Between the two peaks, the first peak is very prominent compared to the second one. The first prominent peak of the signal signifies the GCI of each EGG cycle. Fig. 4d illustrates the DUPEGG signal after suppressing the closing region from the signal. The Figure shows that the secondary peak is emphasised after suppressing the closing region from the signal which actually determines the GOI of each EGG cycle. Figs. 4e and f show the GCI and GOI of the EGG signal detected using the proposed method. Fig. 4g illustrates the corresponding DEGG signal. The positive and negative peaks of the DEGG signal signify the GCI and GOI of the EGG signal.

2.4. Algorithm of the proposed method: In the proposed method, the glottal instants are identified using the unwrapped phase of the EGG signal. The peaks (primary and secondary) of the DUPEGG signal indicate the locations of GCIs and GOIs of the EGG signal. The steps of the proposed phase-based method for detecting the significant glottal instants (GCIs and GOIs) are shown in the algorithm Fig. 5.

2.5. Robustness of the proposed method: The proposed method detects the glottal instants robustly for vulnerable cases of voicing. Fig. 6 illustrates the detection of glottal instants using the proposed method for the low voiced EGG signal. Fig. 6a shows the low voiced EGG signal, After 0.60 s the amplitude of the EGG signal is degraded. In this part of the signal, the cycles of EGG signal are degraded and the knee at GCI and GOI instants are not observable. Fig. 6b represents the DUPEGG signal. The high peak of each cycle of the signal signifies the GCI of the EGG signal. Fig. 6c indicates the DUPEGG signal after suppressing the closing region from the signal. Here, the positive peak of each cycle of the signal represents the GOI of the EGG signal. Fig. 6d illustrates the corresponding DEGG signal. At low voicing part of the signal, the peaks of the signal are not properly observable at GCI and GOI location. This experiment exhibits that the proposed method has robust detection of glottal instants under the vulnerable cases of voicing.

3. Results and discussion: The performance of the proposed method has been measured by identifying the glottal instants from the EGG signal. Existing CoG, VMD and SIGMA methods are used for the comparison with the proposed method. For reference

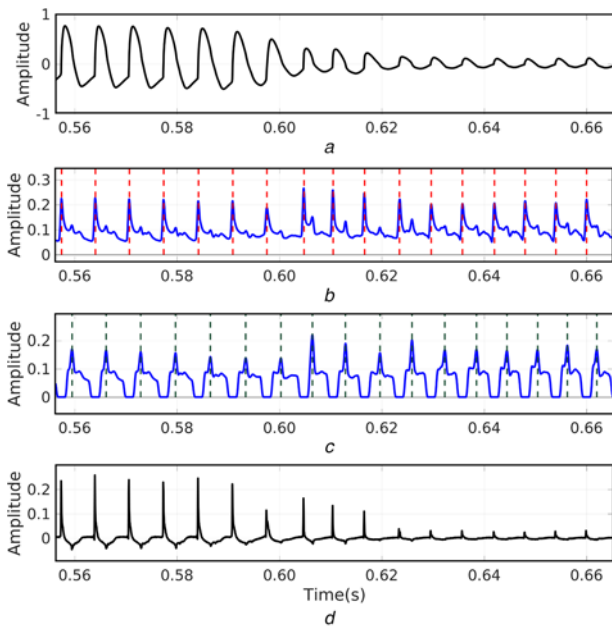


Fig. 6 Illustration of the glottal instants detection using the proposed method for low-voiced EGG signal
a EGG signal with presence of low voicing
b Derivative of the unwrapped phase of EGG signal. The vertical red dashes represent the GCI instants
c Derivative of the unwrapped phase of EGG signal after suppressing the closing regions. The vertical green dashes represent the GOI instants
d Corresponding DEGG signal

purpose, the GCIs and GOIs of these sentences are marked by experts, who are working with speech, EGG and DEGG signals. For each sentence, three signals (speech, EGG and DEGG signals) are provided and asked them to mark the glottal instants by mainly concentrating at the proper discontinuities present in the signals. For evaluation of the proposed method, we have measured the objectives like HR (perfectly one precise instant is identified for each cycle), MR (no instant detected) and FAR (wrongly detected more than one instants within a glottal cycle). The TE (timing error: time difference between the instants marked manually as a ground truth and the instants detected by the method) of the detected glottal instants are calculated with its mean value and the standard deviation (SD).

3.1. Evaluation with CMU-Arctic database: The CMU-Arctic database [24] incorporates speech and corresponding EGG signal of 1132 phonetically balanced English sentences. The database contains three speakers: two male (BDL and JMK) and one female (SLT) speaker. The performance of the proposed method is compared with other existing methods using 300 sentences arbitrarily taken from this database [24]. The database contains a simultaneous recording of EGG and speech signal. The utterances are recorded with the sampling frequency of the signal is 32 KHz [24]. Table 1 shows the performance of the proposed method for detection of glottal instants (GCI and GOI) with comparing the other existing methods. In the case of GCI detection, the proposed method provides 97% detection accuracy, whereas the method provides 96% accuracy in the detection of GOI. The table shows that the performance of the proposed method is slightly better than other existing methods.

3.2. Evaluation with APLAWD database: The APLAWD database contains speech and contemporaneous EGG recordings of five sentences, repeated ten times by five male and five female speakers. The sampling frequency of the recording is kept as

20 KHz with 16-bit resolution [25]. The manually marked dataset is used as ground truth for the detection of the glottal instants. Table 2 shows the overall performance of the proposed and existing methods for the detection of glottal instants. The performances in the detection of glottal instants GCI and GOI are presented separately in the table. Here, the performance of the SIGMA method is better than the other methods. The proposed method provides very close performance to the SIGMA method.

3.3. Evaluation with vulnerable cases of voicing: To evaluate the proposed method, we choose 300 utterances of vulnerable voicing from CMU-Arctic database. Out of these 300 utterances, 120 utterances mentioned as the transition in LMs and 180 utterances correspond to end of voicing. As the EGG signal is degraded in nature for these types of voicing, the accuracy of detection of glottal instants is reduced for each of the methods. Out of these methods, the performance of the proposed method is more desirable than the other existing method. Table 3 shows that

Table 1 Evaluation results of the proposed and other existing methods for detection of glottal instants based on CMU-Arctic database

Instants	Method	HR, %	MR, %	FAR, %	TE in ms	
					Mean	SD
GCI	CoG	90.85	2.18	6.96	0.61	0.31
	SIGMA	96.07	3.57	0.36	0.54	0.29
	VMD	95.86	3.30	0.84	0.44	0.24
	proposed	97.37	2.32	0.31	0.79	0.17
GOI	CoG	87.52	2.88	9.56	0.79	0.44
	SIGMA	95.08	3.65	1.27	0.61	0.34
	VMD	94.83	3.42	1.76	0.52	0.31
	proposed	96.83	2.47	0.71	0.34	0.21

Table 2 Evaluation results of the proposed and other existing methods for detection glottal instants based on APLAWD database

Instants	Method	HR, %	MR, %	FAR, %	TE in ms	
					Mean	SD
GCI	CoG	93.07	1.74	5.19	0.67	0.28
	SIGMA	98.41	1.02	0.57	0.35	0.24
	VMD	97.92	1.41	0.67	0.40	0.21
	proposed	98.23	1.39	0.38	0.33	0.17
GOI	CoG	90.56	2.71	6.72	0.82	0.37
	SIGMA	97.84	1.26	0.91	0.49	0.32
	VMD	97.30	1.76	0.94	0.52	0.28
	proposed	97.49	1.94	0.57	0.44	0.26

Table 3 Evaluation results of the proposed and other existing methods for detection glottal for the vulnerable voices

Instants	Method	HR, %	MR, %	FAR, %	TE in ms	
					Mean	SD
GCI	CoG	81.21	7.02	11.76	0.74	0.47
	SIGMA	86.66	5.34	8.01	0.52	0.33
	VMD	88.90	3.21	7.89	0.55	0.32
	proposed	92.58	3.10	4.32	0.38	0.25
GOI	CoG	73.64	11.93	14.43	0.89	0.56
	SIGMA	80.38	9.27	10.35	0.71	0.43
	VMD	84.33	7.29	8.38	0.63	0.39
	proposed	90.10	4.02	5.88	0.48	0.31

the proposed method provides 92% accuracy for detection of GCIs and provides 90% accuracy for detection of GOIs. The three existing methods provide the accuracy <90% for detection of GCIs and GOIs.

3.4. Evaluation with patients' (vocal folds disorder) database: In this study, we use the database of EGG signals collected from the patients who are suffering from different types of vocal folds disorders. The database consists of both EGG and speech signals (recorded at B. C. Roy Technology hospital, IIT Kharagpur) of the patients who are suffering from pathological disorders in vocal folds. The pathological EGG signals of 72 vfd (vocal folds disorder) patients are analysed in the proposed method. For this database, vowel utterances are considered for recording the EGG signals. The reason for choosing the vowel utterances is mainly for examining the detailed characteristics of vocal folds movements of patients having disorders in their vocal folds. From each patient, five vowels (a, i, e, u and o) utterances are recorded, where in each utterance the specific vowel sound is repeated three times. The patients are asked to utter these vowels with comfortable loudness. The duration of each utterance is around 2–3 s. The sampling frequency of recorded EGG signal is kept as 16 kHz with a resolution of 16 bits. The types of disorders in vocal folds are polyp, nodule, paralysis, ulcer in vocal folds and thicken vocal folds. Table 4 shows the performance of proposed and other existing methods for detecting the GCIs and GOIs for pathological database. The proposed method provides nearly 91% accuracy in detection of GCIs and 88% accuracy in detection of GOIs. The performance of the VMD method is better compared to the other existing methods though the accuracy in detection of glottal instants is <90%. The TE of the proposed method is less compared to the other existing methods. The performance of the proposed method is better to compare with the other existing method. Since the proposed method is based on the signal processing ideas rather than statistical analysis, it is expected that the trend in the values in the tables are not likely to be very different for much larger dataset covering a large number of speakers. Results show that the proposed method can significantly improve the accuracy of the EGG parameter extraction by identifying the glottal instants under different kinds of EGG signal. Based on the detection accuracy, the proposed method is suitable for automated speech technology and vocal fold pathology analysis systems. The performance of the proposed method is slightly better than the rest of the methods in normal voicing. The reason may be due to the fact that the existing methods depend on the strength of the signal. However, the proposed algorithm depends on the phase value of the signal which is independent of the strength of the signal. Moreover, the proposed method performed significantly better in vulnerable case of voicing like the transition in LMs and low voicing.

Table 4 Evaluation results of the proposed and other existing methods for detecting GCIs and GOIs on vocal folds disorder patients' database

Instants	Method	HR, %	MR, %	FAR, %	TE in ms	
					Mean	SD
GCI	CoG	79.85	6.97	13.18	0.97	0.51
	SIGMA	84.91	7.43	7.66	0.60	0.34
	VMD	86.43	7.04	6.21	0.61	0.38
	proposed	90.86	3.47	5.67	0.46	0.26
GOI	CoG	64.89	14.68	20.43	1.16	0.63
	SIGMA	77.74	12.35	9.91	0.77	0.49
	VMD	80.81	8.14	11.05	0.73	0.48
	proposed	87.90	5.05	7.04	0.57	0.36

4. Conclusion: In this paper, an efficient method has been presented based on the phase of the signal for automatically determining the glottal instants from EGG signal including the transition in LMs and low voiced signal. The proposed phase-based method is straightforward and easy to compute. This method draws important characteristics of the phase signal, which has least influenced by strength/amplitude of the EGG signal. The information present in the phase signal mainly depends on the vibratory movement of the vocal folds. Any changes during the movement of the vocal folds are well captured in the phase of the EGG signal. Thus, for some vulnerable cases of voicing where the amplitude or the strength of the EGG signal is disturbed, we can rely on the phase signal for detection of the glottal instants. The most of the existing methods may fail severely in the detection of glottal instants during the vulnerable situations mentioned in the paper, as the strength of EGG signal is significantly low, irregular and disturbed for these types of voicing. Evaluation results demonstrate that the proposed method significantly outperforms the other existing methods under vulnerable cases of voicing and vocal folds pathological signal. Based on the performance in detection of the glottal instants, we can rely on the proposed method for automatic analysis of the glottal activities in robust and accurate way. In future direction, the proposed phase-based method can be useful for extracting the several parameters from the EGG signal which will be applicable for various speech and bio-medical applications.

5. Acknowledgments: The authors thank the authorities of the B C Roy Technology hospital for allowing us to collect data from the patients suffering with vocal disorders. Our sincere thanks to the patients who give their time and patience for collecting the speech and EGG signals.

6 References

- [1] Aicha B., Noureddine E.: 'Electroglottographic measures based on GCI and GOI detection using multiscale product', *Int. J. Comput. Commun. Control*, 2008, **3**, (1), pp. 21–32
- [2] Bachhav P.B., Patil H.A., Patel T.B.: 'A novel filtering based approach for epoch extraction'. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Brisbane, Australia, 2015, pp. 4784–4788
- [3] Kadiri S.R., Yegnanarayana B.: 'Epoch extraction from emotional speech using single frequency filtering approach', *Speech Commun.*, 2017, **86**, pp. 52–63
- [4] Thotappa D., Prasanna S.R.M.: 'Reference and automatic marking of glottal opening instants using egg signal'. Proc. Intl. Conf. on Signal Processing and Communications (SPCOM), Bangalore, India, 2014, pp. 4260–4264
- [5] Kadiri S.R., Yegnanarayana B.: 'Analysis of singing voice for epoch extraction using zero frequency filtering method'. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Brisbane, Australia, 2015, pp. 4260–4264
- [6] Deshpande P., Manikandan M.S.: 'Effective glottal instant detection and electroglottographic parameter extraction for automated voice pathology assessment', *IEEE. J. Biomed. Health. Inform.*, 2018, **22**, (2), pp. 398–408
- [7] Childers D.G., Larar J.: 'Electroglottography for laryngeal function assessment and speech analysis', *IEEE Trans. Biomed. Eng.*, 1984, **31**, (12), pp. 807–817
- [8] Childers D.G., Hicks D.M., Moore G.P., *ET AL.*: 'Electroglottography and vocal fold physiology', *J. Speech, Lang. Hear. Res.*, 1990, **33**, (2), pp. 245–254
- [9] Jyothish Lal G., Gopalakrishnan E.A., Govind D.: 'Accurate estimation of glottal closure instants and glottal opening instants from electroglottographic signal using variational mode decomposition', *Circuits Syst. Signal Process.*, 2017, **37**, pp. 810–830
- [10] Murty K.S.R., Yegnanarayana B.: 'Epoch extraction from speech signals', *J. Speech Lang. Hear. Res.*, 1990, **16**, (8), pp. 245–254
- [11] Koutrouvelis A.I., Kafentzis G.P., Gaubitch N.D., *ET AL.*: 'A fast method for high-resolution voiced/unvoiced detection and glottal closure/opening instant estimation of speech', *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 2016, **24**, (2), pp. 316–328

- [12] Mathur A., Chaudhary N., Upadhyay A., *ET AL.*: 'Detection of glottal closure instants from voiced speech signals using the Fourier-bessel series expansion'. Proc. Intl. Conf. on Communications and Signal Processing, Melmaruvathur, India, 2015, pp. 474–478
- [13] Aneja G., Kadiri S.R., Yegnanarayana B.: 'Detection of glottal closure instants in degraded speech using single frequency filtering analysis', *INTERSPEECH*, 2018, pp. 2300–2304
- [14] Henrich N., dAlessandro C., Doval B., *ET AL.*: 'On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation', *J. Acoust. Soc. Am.*, 2004, **115**, (3), pp. 1321–1332
- [15] Henrich N., Roubeau B., Castellengo M.: 'On the use of electroglottography for characterisation of the laryngeal mechanisms'. Proc. of the Stockholm Music Acoustics Conf., Stockholm, Sweden, 2003, pp. 6–9
- [16] Adiga N., Prasanna S.: 'Detection of glottal activity using different attributes of source information', *IEEE Signal Process. Lett.*, 2015, **22**, (11), pp. 2107–2111
- [17] Brookes M., Naylor P., Gudnason J.: 'A quantitative assessment of group delay methods for identifying glottal closures in voiced speech', *IEEE Trans. Acoust., Speech, Signal Process.*, 2006, **14**, (2), pp. 456–466
- [18] Thomas M.R.P., Naylor P.: 'The Sigma algorithm: a glottal activity detector for electroglottographic signals', *IEEE Trans. Audio, Speech and Lang. Process.*, 2009, **17**, (8), pp. 155–1566
- [19] Roubeau B., Henrich N., Castellengo M.: 'Laryngeal vibratory mechanisms: the notion of vocal register revisited', *J. Voice*, 2009, **23**, (4), pp. 425–438
- [20] Yamauchi A., Yokonishi H., Imagawa H., *ET AL.*: 'Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging', *J. Voice*, 2015, **30**, (2), pp. 205–214
- [21] Sunil Kumar S.B., Mandal T., Sreenivasa Rao K.: 'Robust glottal activity detection using the phase of an electroglottographic signal', *Biomed. Signal Process. Control (Elsevier)*, 2017, **36**, pp. 27–38
- [22] Mandal T., Sreenivasa Rao K.: 'Robust detection of glottal activity using unwrapped phase electroglottographic signal'. Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Calgary, Canada, 2018, pp. 5584–5589
- [23] Oppenheim A.V., Schaffer R.W., Buck J.R.: 'Discrete-time signal processing' (Prentice Hall, Upper Saddle River, NJ, 1999)
- [24] Kominek J., Black A.W.: 'The CMU-Arctic speech databases'. ISCA Speech Synthesis Workshop, Pittsburgh, PA, USA, 2004, pp. 222–224
- [25] Lindsey G., Breen A., Nevard S.: 'SPAR's archivable actual word databases' (Tech. Rep. Univ. College London, London, UK, 1987)