Biomedical Signal Processing and Control 36 (2017) 27–38

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Robust glottal activity detection using the phase of an electroglottographic signal

S.B. Sunil Kumar[a], Tanumay Mandal[b,*], K. Sreenivasa Rao[a]

[a] Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, 721302, India
[b] Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, 721302, India

ARTICLE INFO

Article history:
Received 19 April 2016
Received in revised form 5 January 2017
Accepted 22 March 2017

Keywords:
Electroglottography
EGG signal
Phase of EGG signal
Glottal closure instant
Glottal opening instant
Laryngeal transition

ABSTRACT

Electroglottography (EGG) is a non-invasive technique of studying the glottal activity. The phase of EGG signal has not attracted much attention, although it is robust and capable of revealing dynamics of glottal activity. The main objective of this work is to develop a robust and accurate technique for analyzing the glottal activity by exploiting the phase of EGG signal. Using the phase of EGG signal, accurate and robust methods are proposed to detect glottal closure instant (GCI) and glottal opening instant (GOI) within a glottal cycle. The glottal activity during close and open phases of a glottal cycle can be precisely analysed by using GCIs and GOIs respectively. Robustness of the proposed method for detection of GCIs and GOIs is also evaluated in vulnerable cases such as transition in laryngeal mechanism, weak voicing and voice transition. The accuracy of the proposed method is evaluated using manually labeled EGG signals chosen from CMU Arctic database, and the performance of the proposed method is also compared with the state-of-the-art singularity in EGG by multiscale analysis (SIGMA) method. Finally, the performance of the proposed and SIGMA methods is also evaluated on EGG signals recorded from the patients having vocal folds disorders.

1. Introduction

Electroglottography (EGG) is a well-known approach to investigate the activity of the vocal folds in a non-invasive way [1]. The dynamics in the vocal folds vibration during speech production is well captured by EGG device in the form of variation of electrical admittance between the electrodes. The admittance between the electrodes is proportional to the contact area between the vocal folds. When vocal folds are in closed position, the admittance between the electrodes of EGG device is maximum, whereas it is minimum, when vocal folds are in opened position. The time instant, when the vocal folds start attempt to close is known as glottal closing instant (GCI) and the time instant, when the vocal folds start attempt to open or separate from each other is known as glottal opening instant (GOI). Detection of GCIs and GOIs from EGG signal is important for several speech and laryngeal pathological applications. Accurate and robust detection of GCIs and GOIs can improve the performance of several speech applications such as glottal-synchronous speech synthesis [2], pitch tracking [3], speech prosody modification [4] and speech dereverberation [5]. GCIs and GOIs are also used to categorize various types of vocal folds disorders by utilizing the specific information in terms of glottal pulse parameters such as duration of open and close phases of the glottal cycle, shape of glottal volume velocity and strength of excitation at GCIs and GOIs [6,7]. As the regions around GCIs have high signal to noise ratio, processing the speech anchored around these instants will be preferred for most of the speech tasks [8–10]. Therefore, accurate detection of GCIs and GOIs will play an important role in most of the speech tasks [10].

Childers et al., showed that the parameters of EGG signal can be used as an aid to diagnose laryngeal dysfunction [7]. Further, parameters of the glottal cycle are used in the analysis of speech pathological conditions such as dysphonia [11]. Hence, robust and accurate detection of GCIs and GOIs will benefit a good spectrum of speech processing and bio-medical applications. Childers et al., have studied glottal activity using EGG, Differentiated EGG (DEGG) and ultra high-speed cinematography [12–14,7,15]. From the above studies, they showed that instants corresponding to positive and negative peaks in DEGG signal can be used to determine approximate GCIs and GOIs, respectively. As the closing phase of the glottal cycle is fast and abrupt in nature, EGG signal shows sudden rate of increase in admittance. Hence, a very strong and precise positive peak can be expected in DEGG signal at GCI location. This leads

* Corresponding author.
E-mail addresses: tanum.dets@gmail.com (T. Mandal), ksrao@iitkgp.ac.in (K. Sreenivasa Rao).

http://dx.doi.org/10.1016/j.bspc.2017.03.007
1746-8094/

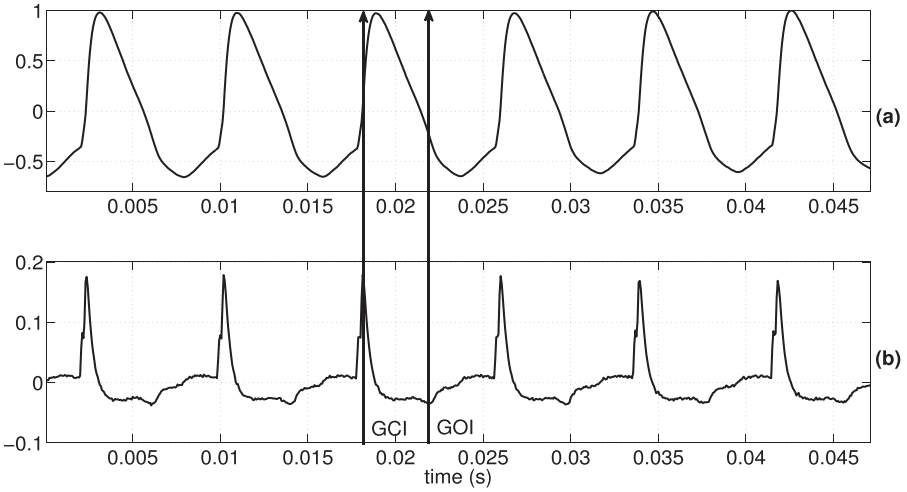

Fig. 1. Illustration of detection of glottal closure instants (GCIs) and glottal opening instants (GOIs) using positive and negative peaks in DEGG signal, respectively. (a) EGG signal over six glottal cycles and (b) the corresponding DEGG signal.

to accurate detection of GCI. Slow paced nature of opening phase yields a very weak and broad negative peak in DEGG signal, which creates uncertainty in making the decision of the exact instant of glottal opening [1]. Fig. 1 shows the GCIs and GOIs correspond to positive and negative peaks of DEGG signal, respectively.

Fig. 1(a) and (b) show EGG and its corresponding DEGG signals, respectively. In literature, most of the methods for detection of GCI and GOI have exploited either EGG or DEGG signals. Rothenberg proposed a threshold based method for detecting GCIs and GOIs [16]. In this method, a percentage of threshold amplitude between the minimum and the maximum of the EGG signal over a glottal period is chosen. Time instants where these thresholds intersect the EGG signal are marked as GCIs and GOIs. Such threshold-based methods are imprecise and inaccurate [17]. In addition to time domain approaches mentioned above, methods based on wavelet transform have been explored for accurate detection of GCIs [18–20]. Further the Lines of Maximum Amplitudes (LOMA) of the wavelet transform have been investigated for accurate and reliable detection of GCIs [21,22]. In 2009, Thomas and Naylor, have proposed Singularity in EGG by Multiscale Analysis (SIGMA) algorithm for accurate detection of GCIs and GOIs, and the detected instants are used for the analysis of glottal activity [11]. In this paper, authors first detected the significant changes in EGG signal by using multiscale product of 3 dyadic scales. Then group delay function is used to determine the peaks in multiscale product and Gaussian mixture models are used to remove the spurious instants. Even though SIGMA method demonstrates impressive performance, it needs to tune several parameters at various stages to provide optimal performance. In the recent investigation, Herbst *et al.* have noticed that DEGG peaks do not always coincide with the events of glottal closure and opening. In this work, authors have examined EGG and super-high-speed video recordings for analyzing the activity of vocal folds within a glottal cycle [23]. There have been several parallel attempts to detect GCIs and GOIs using speech signals, in addition to EGG signals [24–30]. A detailed survey of GCIs and GOIs detection methods and their performance comparison are provided in [31,32].

The state-of-art methods detect GCIs and GOIs based on EGG signal, and they provide accurate results during normal voicing conditions. However, most of the existing methods may fail to detect GCIs and GOIs during the laryngeal mechanisms(LM) where the DEGG signal is showing double peak at closing (or opening) phase of glottal cycle [23]. In LM transition region, the double peak in DEGG signal due to the zipper like closing (or opening) of the vocal folds through anterior-posterior dimension. The closing (or opening) of vocal folds start at the anterior (or posterior) position and proceed up to the posterior (or anterior) position and makes a full closure (or open) of vocal folds with a certain amount of temporal delay. In case of voice transition (voicing from normal to non-voicing i.e., at the end of sentence or phrase), there will be a gradual decay in the strength of EGG signal. These voice transitions, usually occur where the speaker is about to take a long breath or introduce a pause in between the phrases. The decay in EGG signal strength at the syllable or word boundary is viewed as weak voicing in this work. Therefore, the weak voicing is observed within the words and phrases, and voice transition is observed at the end of a phrase or sentence. Hence, detection of precise GCI and GOI locations is difficult during the transition of LM, weak voicing and voice transition by the existing methods.

In case of vocal disorder patients, the EGG signal is irregular and noisy compared to the modal one. There are various types of vocal folds disorder namely, vocal folds thickening, nodule, polyp, ulcer and paralysis. Vocal folds thickening occurs due to uttering of loud or shouted (high intensity) speech for a long time span, and the person having this disorder has the hoarseness in his/her speech. The symptoms of this disorder can be observed in the characteristics of voice such as variation in pitch (F0), and the person having the disorder, feels tiredness and feeling of having to clear his/her throat, after speaking over a long period of time. Nodule and polyp can be observed as extra-growth results on inner and outer layer tissues of the vocal folds, respectively due to vocal abuse or long-term exposure to irritants. The symptoms for both these disorders appear to be similar. The patients having these disorders have hoarseness, low-pitched voice, breathy voice and sometime singers may notice a loss of vocal range [42,43]. Ulcer over the vocal folds is mainly due to acid reflux and bad habits such as frequent smoking, consumption of alcohol and tobacco [43]. The patient having this disorder feels pain while speaking and the person may not speak at all in case if the ulcer is in advanced stage. The vocal disorder related to paralysis of vocal folds is mainly observed in stroke patients and the patients who have surgical trauma (most often from thyroid surgery) where one or both vocal folds are paralyzed, and their movement is restricted based on severity of the disorder. With this effect, their speech is either completely or partially affected and the person feels discomfort from the strain of trying to move paralyzed cords [43–45]. In these vocal folds disorder cases, the existing methods give the poor accuracy of detection of GCIs and GOIs locations due to the nature of the EGG signal.