Interspeech 2018
2-6 September 2018, Hyderabad

Classification of disorders in vocal folds using Electroglottographic Signal

Tanumay Mandal, K. Sreenivasa Rao and Sanjay Kumar Gupta

Indian Institute of Technology Kharagpur, India
tanumay.mandal10@iitkgp.ac.in, ksrao@iitkgp.ac.in, entdr.sanjay@gmail.com

Abstract

The main objective of this paper is to accurately classify the pathological voice based on the disorders in vocal folds. For this purpose, we have explored the phase of Electroglottographic (EGG) signal which carries significant information related to characteristics of vocal folds. Four important parameters, namely, close quotient, open quotient, average pitch period and jitter computed from the phase of the EGG signal have been explored for discriminating the patients based on the disorder in their vocal folds. These parameters have been used for classification of three types of vocal folds disorders: vocal nodules, vocal polyps and laryngitis. In this study, we have used the EGG signals of seventy-nine patients having disorders in vocal folds, collected from hospital. The database contains the simultaneous recording of speech and EGG signals of four vowel ('a','e','o','u') utterances. The experimental result shows that the proposed features extracted from phase, performed well in classification of patients according to the disorder in their vocal folds.

Index Terms: Electroglottography, vocal folds disorders, phase signal.

1. Introduction

The lateral vibration of the vocal folds modulates the air from the lungs and generates the carrier signal of speech. The movement of the vocal folds is quasi-periodic for normal or modal voice. Disordered voice production is mainly due to irregularities or asymmetries of the vocal folds movement leading to disturbances in the voice signal [1]. Physiological alterations of the vocal folds cause unhealthy patterns of vocal folds vibration and decrease in patient speech signal quality known as voice pathologies. Consequently, the detection of incipient damages to the folds is useful for improving the prognosis, treatment and care of such pathologies. Physicians often use invasive techniques, such as endoscopy, laryngostroboscopy or surgical microlaryngoscopy to diagnose symptoms of voice disorders. In state-of-art, researchers used the invasive technique to visualize the vibration of the vocal folds [2, 3, 4, 5]. It is, however, possible to identify disorders using certain features of speech signal in a non-invasive way [6].

In the proposed method, Electroglottography (EGG) is used to extract the parameters for classifying the disorders in vocal folds. EGG is a well-known approach to investigate the activity of the vocal folds in a non-invasive way [7]. In this technique, a high frequency ($f \simeq 1MHz$) modulated current is sent through the subject's thyroid cartilage, by a pair of electrodes held in contact with the skin on either side of the subject's neck [7]. The dynamics of 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 open 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) [8, 9]. Accurate detection of glottal instants will help to extract the parameters robustly. Therefore, in this work, we extract the phase of EGG signal for detection of GCIs and GOIs. The detection of glottal instants using phase information of EGG signal will be highly robust and accurate [10]. The parameters are extracted using these glottal instants. In the proposed work, we classified three types of disorders in vocal folds , namely, vocal nodule, vocal polyp and laryngitis. Laryngitis occurs due to loud utterance of 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 patient's variation of pitch (F0), rough voice quality, tiredness, and temptation to clear his/her throat, after speaking for a long period of time. Nodule and polyp can be observed as extra-growth results on inner and outer layers tissue 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, sometimes, loss of vocal range of a singer.

The organization of this paper is as follows: Section 2 discusses the proposed method to detect the glottal instants and extract the parameters using the glottal instants. The performance of the proposed method is evaluated in Section 3. Finally, Section 4 summarize the contributions of this work.

2. Proposed Method

2.1. Pre-Processing

In EGG recordings, the real-valued discrete time EGG signal can de 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 EGG signals recorded from vocal folds disorders patients have more irregularities and noisy than normal EGG signals. The pre-processing technique helps to decouple the original EGG waveform from low-frequency noise like components and segregate the voiced segments from the silence/unvoiced segments of the captured EGG waveform. At the time of recording, low-frequency components are coupled with the recorded EGG waveform due to swallowing and breathing by the speaker, miss contact of the electrodes with the skin, and non-uniform movement of vocal folds. Presence of low-frequency components delivers uneven amplitude envelope in the EGG signal which influences the detection accuracy of glottal instants. Preprocessing techniques are used to remove those anomalies from the recorded EGG signal. High pass filter with cut-off frequency of 30 Hz has been used for removing the low-frequency

10.21437/Interspeech.2018-1967

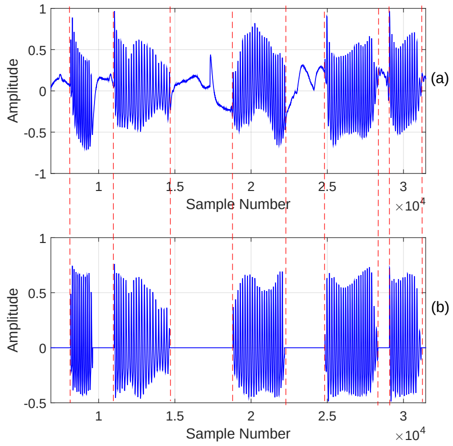

Figure 1: *Illustration of captured EGG signal and the pre-processed EGG signal. (a) Recorded EGG signal (b) Pre-processed EGG signal*

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.

2.2. Unwrapped phase extraction from EGG Signal

The analytic signal $e_a(n)$ corresponding to the EGG signal $e(n)$ is given by [11]

$$e_a(n) = e(n) + je_h(n) \tag{1}$$

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:

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

$$\phi(n) = \tan^{-1}\left(\frac{e_h(n)}{e(n)}\right) \tag{3}$$

Figures 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

$$\Delta\Phi(n) = \Phi(n) - \Phi(n-1) \tag{4}$$

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

$$\Phi_u(n) = \Delta\Phi(n) + 2\pi \tag{5}$$

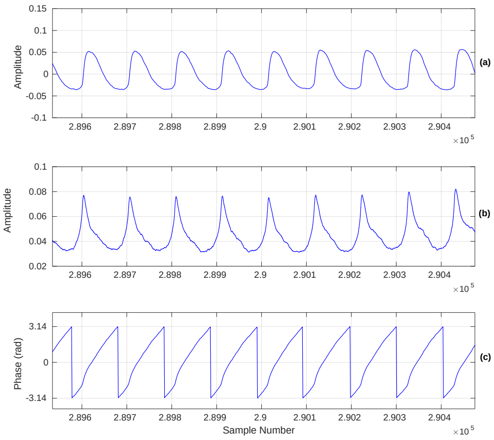

Figure 2: *Instantaneous amplitude envelope and phase of EGG signal computed over ten cycles of EGG signal. (a) EGG signal, its (b) instantaneous amplitude envelope and (c) phase of the signal.*

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. Figures 3(b) and 3(c) show

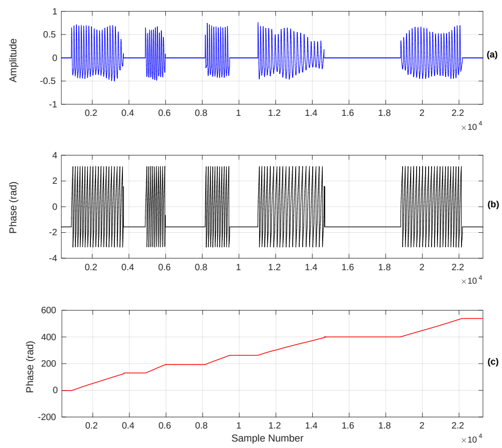

Figure 3: *Phase Extraction from the analytic signal (a) EGG Signal (b) Wrapped Phase (c) Unwrapped Phase*

the corresponding wrapped and unwrapped phase of the EGG signal. In wrapped phase, the phase signal is wrapped according to the period of glottal cycle where, the unwrapped phase is monotonically increasing as the number of glottal cycles are increased.

2.3. Detection of glottal instants

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 major 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 within the signal.

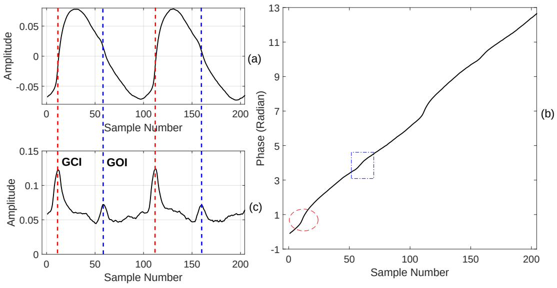

Figure 4: *Illustration of unwrapped phase and differentiated unwrapped phase (a) EGG signal (two cycles) (b) Unwrapped phase of the EGG signal (c) Differentiated unwrapped phase of the EGG signal*

Figures 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 each 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 figures 4(a) and 4(c), the vertical dash line (in red color) represents the GCI location for each glottal cycle. The high 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.

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 and smooth 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. The peaks, present at the GCIs location, are clipped off and emphasize the secondary peak of 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. Now we can visualize the precise peak present 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.

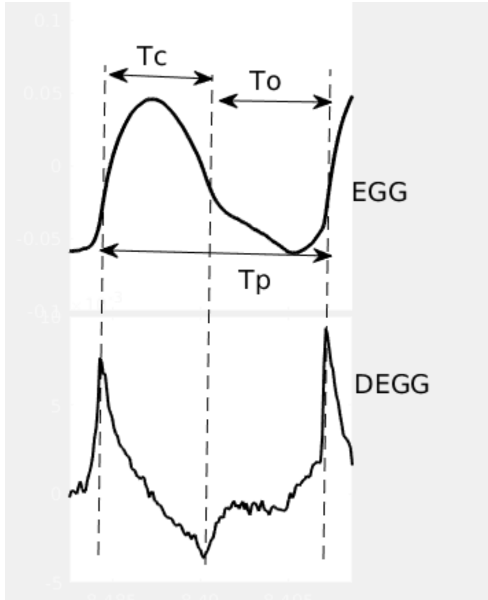

Figure 5: *Illustration of EGG and DEGG signals with time stamps*

2.4. Extraction of parameters

The four predefined parameters have been extracted using the glottal instants of the EGG signal. The glottal instants (GCIs and GOIs) are extracted using the phase of EGG signal. The figure 5 shows the EGG signal (one cycle) and corresponding DEGG signal. The positive peak of the DEGG signal defines the GCI location whereas the negative peak of the DEGG signal defines the GOI location of the EGG signal. The time duration " Tc " define the close period of one glottal cycle. The close period of the glottal cycle starts at the GCI and it continues upto the GOI. Similarly, time duration "To " defines the open time of one glottal cycle. It starts at the GOI and ends at the GCI of the next glottal cycle. The detail information of these parameters are discussed below:

- *Pitch Period :* Pitch period (PP) is defined by the time to complete one full cycle of EGG signal. In figure 5, The time duration " Tp " which starts from GCI and end at the next GCI of the EGG signal, defines the PP of the EGG signal.
- *Jitter :* Perturbation in pitch period of the glottal cycle is known as jitter. Here, the jitter ratio is measured by calculating the absolute value of the difference in time between consecutive glottal cycles over the average glottal period. The jitter ratio (JR) is calculated as

$$JR = |Tp(n) - Tp(n+1)|/Tp(avg) \quad (6)$$

- *Close Quotient :* The close quotient (CQ) is the ratio of the close time to the pitch period of the glottal cycle. The close time taken to the vocal folds to reach fully closed position from start of closing of the vocal folds. The mathematical expressions of the CQ is

$$CQ = Tc/Tp; \quad (7)$$

- *Speed Quotient :* The speed quotient (SQ) is the ratio of open and close time of each glottal cycle. The open time defined as the time taken to the vocal folds to reach fully abduct position from start of separation of vocal folds. The SQ can be expressed as

$$SQ = To/Tc; \tag{8}$$

3. Evaluation of results

The proposed method is evaluated on EGG signals collected from the patients who have the disorders in vocal folds. For carrying out this study, we have prepared a database named IITKGP Vocal Folds Disorder Database (IITKGP_VFDD). The database contains the simultaneous recording of EGG and speech signals collected from the vocal folds disordered patients of B. C. Roy Technology Hospital, IIT Kharagpur. In this database, the patients are asked to utter four vowels (‘a’,‘e’,‘o’,‘u’), each vowel for three times. The sampling rate is kept 16 KHz with 16 bit resolution.

In the proposed work, we have used EGG signal of seventy-nine vocal folds disorders patients for evaluation. Among them, twenty-five patients have nodule in vocal folds, twenty-two patients have polyp in vocal fold and remaining thirty-two patients have laryngitis. The parameters extracted from the phase of EGG signal are used to categorize the disorders in vocal folds in robust way. The pitch-period of a normal person is varying according to the age and gender. The pitch-period of a vocal folds disorder patient changes significantly compared to the normal person. By measuring the pitch period, a person can be distinguished either he/she has some disorders in vocal folds or not. But, the type of disorder in vocal folds is very difficult to categorize using the pitch-period.

Table 1: *Statistical Evaluation on three types disorders in vocal folds*

Disorder Name	No. of Patients	Jitter	CQ	SQ
Vocal Nodule	25	0.076±0.004	0.431±0.11	1.32±0.628
Vocal Polyp	22	0.061±0.005	0.592±0.077	0.687±0.197
Laryngitis	32	0.045±0.002	0.534±0.051	0.872±0.171

The parameters like jitter, CQ, SQ are capable to distinguish various types of disorders in vocal folds. Table 1 shows the statistical result of three types vocal folds disorder. The mean value and the standard deviation of each parameter is extracted for three types of vocal folds disorder. Here, the extracted parameters show the threshold based classification of three types of vocal folds disorders. The parameters are taken as a average value of the patients for a particular disorder in vocal folds. In table 1, it is shown that the jitter is more for vocal nodule patients compared to the other two disorders. CQ is more for polyp patients as the close time of vocal fold is more for polyp patients. The SQ is high for vocal nodule patients compared to the vocal polyp and laryngitis patients. Figure 6 represents the graphical view of classification of vocal folds disorders. The parameters are presented in x-axis and the average value for each parameter for the particular disorders are presented in y-axis.

Seventy two segments (two patients from each type of vocal folds disorder X four vowels for each patient X three times utterances for each vowel) of EGG signal are used as a test signal for evaluation of the proposed work. Table 2 shows the confusion matrix for classification of three types vocal folds disorder using the extracted parameters.

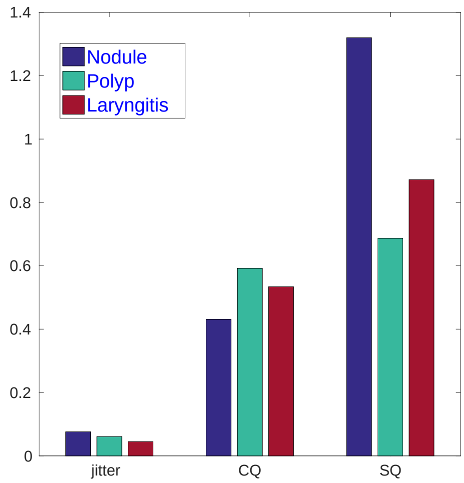

Figure 6: *Graphical representation of classification of disorders in vocal folds*

Table 2: *Confusion matrix for classification of vocal folds disorders*

	Vocal Nodule (%)	Vocal Polyp (%)	Laryngitis (%)
Vocal Nodule	87.57	4.17	8.34
Vocal Polyp	4.17	83.40	12.51
Laryngitis	8.34	12.51	79.23

The parameters extracted from each segment are compared with the mean value of each parameter of three types of disorders in table 1. The result shows that the proposed method classified these three types vocal folds disorder based on the extracted features with good accuracy.

4. Conclusions

A simple method is proposed for classifying the disorders in vocal folds. The parameters which are extracted from the EGG signal, are useful to classify the disorders. The threshold of the each parameters distinguish the disorders from each other though the value of parameters can be changed according to the size and position of the vocal nodule and vocal polyp. The results show that it is possible to distinguish one kind of vocal folds disorder from others by measuring the threshold value of extracted parameters for each kind of diseases but more strict control means of data acquisition should be devised and some machine learning technique requires more evaluation before clinical applications.

5. Acknowledgments

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

6. References

[1] 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.

[2] J. Unger, M. Schuster, D. J. Hecker, B. Schick, and J. Lohschelle, "Voice pathology classification based on high-speed videoendoscopy," *Engineering in Medicine and Biology Society (EMBC)*, vol. 35, pp. 7360–7363, 2013.

[3] D. Panek, A. Skalski, T. Zielinski, and D. Deliyski, "A multiscale product approach for an automatic classification of voice disorders from endoscopic high-speed videos," *Engineering in Medicine and Biology Society (EMBC)*, vol. 37, pp. 735–738, 2015.

[4] H. I. Turkmen, M. E. Karsligil, and I. Kocak, "Classification of laryngeal disorders based on shape and vascular defects of vocal folds," *Computers in Biology and Medicine*, vol. 62, pp. 76–85, 2015.

[5] Jung-HaeCho, Yong-SugChoi, Young-HoonJoo, Young-HakPark, and Dong-IlSun, "Clinical significance of contralateral reactive lesion in vocal fold polyp and cyst," *Journal of voice*, vol. 32, no. 1, pp. 109–115, 2018.

[6] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, "Detection of pathological voice using cepstrum vectors: A deep learning approach," *Journal of voice*, 2018.

[7] Baken and R. J, "Electroglottography," *Journal of Voice*, vol. 6, no. 2, pp. 98–110, 1992.

[8] 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.

[9] 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.

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

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