

Review of Egg and Speech Processing Techniques for Glottal Activity Detection

B.Gayathri¹ and Dr. Malini Mudigonda²

^{1,2}Department of Biomedical Engineering, University College of Engineering, Osmania University, Hyderabad, India
Email: gayathri.b@uceou.edu, malini.m@uceou.edu

Abstract—Glottal instants namely GCIs and GOIs are useful in a wide variety of speech processing and biomedical applications. This paper presents the recent developments in the methodologies for glottal activity detection using EGG and speech signals. The databases used by the researchers are APLAWD, SAM, KEELE and BDL. The methodologies are compared in terms of identification rate, miss rate, and false alarm rate and identification accuracy. The applications in which these methods are employed are also presented. It is observed that many researches are conducted to identify the glottal activity based on the speech signal rather than the EGG signal.

Index Terms— Glottal activity, GCI, GOI, EGG, speech signal.

I. INTRODUCTION

Laryngeal physiology is an area of research to examine the functioning of larynx in the human respiratory system. Larynx contains vocal cords or folds which are essential for human speech and hence called as the voice box. Vocal folds are primarily responsible for respiration, swallowing, phonation and speech. The vibration of vocal folds at the glottis region produces speech signal by the excitation of the vocal tract. The ligaments in the vocal fold form a constriction on the top of trachea when the excitation is caused by the vocal folds. As the air is expelled from the lungs, partial vacuum cause pressure on the vocal folds and makes them snap shut and disrupts the air flow. This is known as glottal activity (GA). Irregular flow of air flow through the glottis leads to turbulence (non-GA). The strength of excitation during the glottal activity is determined by the rate of closure of vocal folds in glottal cycle. The glottal closure instant (GCI) occurs when the glottis is closed. The muscle causes the folds to reopen and it causes glottal opening instant (GOI). This process repeats periodically and constitutes one glottal cycle.

Electroglottography (EGG), also known as laryngograph is a non-invasive technique which measures the impedance of the vocal folds contact area by placing two electrodes across the neck. It tracks the glottal vibrations and records the electrical conductance of the glottis with respect to speech. The waveform thus obtained is known as EGG signal. Identifying the GCIs and GOIs in the EGG signal represents the glottal activity. These particular events (GCIs and GOIs) correspond to high energy and low energy in the glottal signal. The need of the hour is the accurate identification of GCIs and GOIs. This can be done either using the EGG signal or the speech signal. GCIs and GOIs are useful in wide variety of speech processing and biomedical applications. These applications are pitch tracking, prosodic speech modification, , feature extraction for speaker identification, pathological speech, speech synthesis, voice quality enhancement, speech coding and

transmission, identity claim of a speaker, phonetic factors, emotional expression, linguistic feature, psychology and many more. Further research in this area also helps in differentiating voice pathologies like dysarthria, dysphonia, dysphonia and Parkinson's etc with a normal or healthy speech.

II. AIM

The purpose/aim of this review paper is to evaluate the current or existing methods of identification of GCIs and GOIs using electroglottograph signal or a speech signal and its applications.

III. CURRENT RESEARCH ON ESTIMATION OF GCIS AND GOIS

Most of the current research to estimate the glottal activity is based either on EGG signal or on speech signal. Different methods are implemented on the EGG ([1] to [14]) and speech signal ([15] to [36]) to identify accurate peaks of GCI and GOI accurately. The literature survey exposed the various challenges in implementing the algorithms. While some have missed some events of the activity or have duplicates of the same, many algorithms cannot be implemented directly on the EGG signal to clearly distinguish GCIs and GOIS. Glottal activity detection, especially GOIs, from the speech signal is also a challenging task. Overview of recent developments is presented in Table I.

TABLE I. RECENT DEVELOPMENTS TO DETECT GLOTTAL ACTIVITY

Year	Authors	Method Used	Type of signal/ database used
1979	Tirupattur V. Ananthapadmanabha and Yegnanarayana	LP Residual Hilbert Envelope	speech signal
1989	Douglas O'Shaughnessy, Yan Ming Cheng	LP Residual Hilbert Envelope	EGG Signal
1990	K. S. R. Murty and B. Yegnanarayana	Zero Frequency Resonator	Speech signal
1995	Roe1 Smits and B. Yegnanarayana	Group Delay	Speech Signal
1996	L. Janer, J.J. Bonet, E. Lleida-Solano	Wavelet Transform	speech signal
1999	Vu Ngoc Tuan & Christophe d'Alessandro	Line Of Maximum Amplitude (LOMA)	Speech signal
2002	AnastasisKounoudes, Patrick A. Naylor, Mike Brookes	DYPSA	EGG signal
2006	Mike Brookes, Patrick A. Naylor, and Jon Gudnason	Group Delay	Voice speech
2006	Anand Joseph M., Guruprasad S., Yegnanarayana B.	Group Delay	Speecg signal
2007	K. S. Rao, S. R. M. Prasanna, and B. Yegnanarayana	HE(Hilbert Envelope) Using Group Delay	Speech signal
2007	Mark R. P. Thomas, Nikolay D. Gaubitch and Patrick A. Naylor	DYPSA	Speech signal
2007	Patrick A. Naylor, AnastasisKounoudes, Jon Gudnason, and Mike Brookes	DYPSA	Speech signal
2008	B. Aicha and E. Noureddine,	Cubic Root Of Multiscale Product(MP)	EGG
2008	WafaSaïdi, AïchaBouzid, NoureddineEllouze	MP And DYPSA	Speech signal
2008	Mark R. P. Thomas, Jon Gudnason and Patrick A. Naylor	DYPSA	Speech Signal
2009	K. Sri Rama Murty, B Yegnanarayana and M. Anand Joseph	Zero Frequency Filtering(ZFF)	Speech signal
2009	N. Dhananjaya, B. Yegnanarayana,	ZFF	Speech Signal
2009	M. R. P. Thomas and P. Naylor	SIGMA	EGG
2011	Christophe D'alessandro And Nicolas Sturm	LOMA	Speech Signal
2012	Mark R. P. Thomas, Jon Gudnason, and Patrick A. Naylor	YAGA	Speech signals
2013	Ramesh K, S. R. M. Prasanna and D. Govind	ZFF & Hilbert Envelope	EGG
2014	D. Thotappa and S. R. M. Prasanna	Multiscale product	Speech signal
2015	A.I. Koutrouvelis, G.P. Kafentzis, N.D. Gaubitch, R. Heusdens	GEFBA	Speech Signal
2015	P. B. Bachhav, H. A. Patil, and T. B. Patel	Novel Filtering Method	EGG
2017	S.B. Sunil Kumar a, Tanumay Mandal b, K. Sreenivasa Rao	Phase Of Egg	EGG
2017	S. R. Kadirir and B. Yegnanarayana	Single Frequency Filtering(SFF)	Speech signal
2018	Tanumay Mandal , K. Sreenivasa Rao	Unwrapped Phase Of Egg Signal	EGG

IV. METHODOLOGIES TO DETECT GCI AND GOI BASED ON EGG SIGNAL

Many algorithms are proposed to detect the glottal instants from the EGG signal. The detailed review of the algorithms is presented in the following sections.

A. Threshold based method

An easy and simple way to analyze an EGG signal is by comparing the spikes of dEGG signal with the threshold values. Some researchers used 50% [1], while some others used 3/7th of the maximum EGG amplitude as threshold to find GCI and GOI [2]. Huckvale employed HQT_x (High Quality Time of excitation) and TXGEN (Time of eXcitation GENerator) methods [3] to apply a threshold that varies dynamically with the signal and the instantaneous gradient. While HQT_x can detect only GCI, TXGEN can detect both GCI and GOI. In these method the extrema of dEGG signal when multiplied by a constant value are fixed as the high and low thresholds. If the signal crosses both the thresholds within a period of time, a GCI instant is flagged. In TXGEN, the GOI is flagged at the point on EGG when its amplitude equals the amplitude at the preceding GCI. However, the drawbacks of the threshold based methods are the poor accuracy for GCI and inability to detect GOIs.

B. Group delay Method

Different group delay (overall transmission delay for a signal) methods namely phase characteristics, strength of excitation, energy weighted measure are used to determine the glottal instants. Phase characteristic measure uses wrapped minimum phase signals [4]. To unwrap the signal, group delay values (continuous and differ by minimal amount) are calculated at each frequency. Average slope of the phase spectrum is obtained through linear regression method. The instants that make positive and negative zero crossings in the phase slope function are identified as the GCIs and GOIs respectively. Another method [5], focuses on phase characteristics of the excitation signal. The computation of average slope of the unwrapped phase spectrum is directly obtained by calculating group delay function from the windowed signal. The analysis window is moved by one sample at a time to get the phase slope function. The positive zero-crossings represent the instants of significant excitation. Another method [6], proposes the strength of excitation as a measure of glottal excitation instants, determined by the energy of the residual error signal surrounding the excitation. The strength of excitation is calculated by Frobenius norm. The Frobenius norm is computed at each instant of the sample and the locations of peaks in the plot determine the desired instants. The last method namely the energy weighted measure of the group delay function [7] also identifies the glottal instants as the negative zero crossings obtained from (linear prediction)LP residuals. Compared to other group delay characteristics, the results are found to be more accurate in this method.

C. Multiscale Product Method

The signal singularities (impulse/ abrupt changes of a signal) are efficiently detected by wavelet coefficients. Singularities of EGG signal represent GCI at modulus maxima and GOI at modulus minima [8]. Accuracy of this method is poor as it cannot detect some singularities for small scales of wavelet transform. Hence multiscale analysis [9] of the wavelet transform helps in improving edge detection. Product of adjacent scales (dyadic scales), improves the peaks and sharpens the edges. For further peak enhancement, we use cubic root amplitude product of the dyadic scales.

D. SIGMA (Singularity In EGG by Multiscale Analysis) Method

This algorithm [11] uses group delay method instead of using cubic root amplitude product [9] for better peak detection. This approach further enhances the detection of glottal instants by using Gaussian mixture to eliminate the unlikely features of noise and the incorrect estimates. However, detection of GOIs includes further processing. The method is found to be more appropriate than the threshold methods ([1]-[3]) (Figure.1).

E. Novel filtering Method

Here the speech signal is directly taken from the EGG signal [12]. The method claims to detect epochs in a speech signal. The epoch is defined as the GCI of the vocal tract system. In the proposed method, the signal is positively clipped and then inverted. The resultant signal is passed through third order Butterworth filter. The delay introduced by the filter in the signal is adjusted, and the group delay is calculated at cutoff frequency. The resultant is compared with dEGG as reference signal and found that the negative peaks coincide with peaks of filter output which closely indicates GCI. The disadvantage of the system is it cannot detect GOIs.

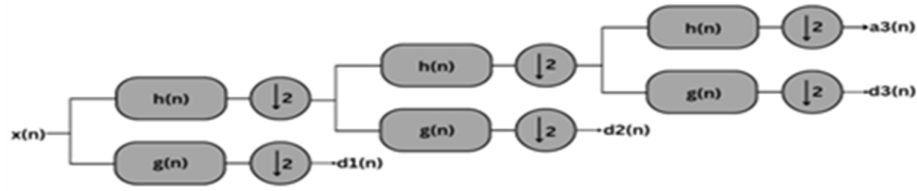


Figure 1.SIGMA Algorithm

F. Phase of EGG signal

As the EGG signal is time varying in amplitude, this method [13] studies the glottal activity based on the phase and frequency of EGG signal. The phase of the EGG signal preserves its characteristics, as the sensitivity of phase depends more on the movement of vocal folds than on the signal amplitude. Hilbert transform is used to detect instantaneous phase of an EGG signal. The sudden rise of the EGG's phase signal represents a GCI. And the phase of EGG signal, when passed through the low pass filter represents GOIs at the positive-to-negative zero-crossings.

G. Unwrapped phase of EGG signal

Hilbert transform is implemented on the EGG signal to detect voiced and non voiced regions [14]. The phase signal is uniformly wrapped throughout the voiced regions and non-uniformly wrapped in non-voiced regions. To identify the closing and opening instants, phase unwrapping is required. The unwrapped phase function can add or subtract any integer multiple to obtain clear discontinuity. The discontinuities represent the GCI of the EGG signal. The peak of GCI is clipped off from the first order derivative of the unwrapped phase of EGG signal and the second peak thus indicates GOI instant.

V. METHODOLOGIES FOR GCI AND GOI DETECTION FROM SPEECH SIGNAL

Though EGG signal effectively tracks the glottal activity, it fails to analyze the functioning of the glottis as it works with the impedance between the vocal folds rather than on the air flow. Hence the glottal activity detection in speech signals is of higher importance as it is oriented towards real time speech processing applications. The detailed review of the algorithms is presented in the following sections.

A. Group delay method [4]

As discussed in section 3.1.2, can also be applied to speech signals apart from EEG signals. Block diagram of group delay method is shown in Figure 3.

B. The LOMA method

The method [15] is proposed to detect glottal GCIs using time frame analysis. Dyadic wavelet transform is applied on speech signal and amplitude maxima are detected. It is observed that the glottal signal forms lines of maximum amplitude [LOMA] in the time domain. A tree like pattern is formed from these lines for each pitch period and the top of the tree is considered as a GCI instant. An improved algorithm [16] performs better by introducing backtracking of the algorithm. LOMA contains more information than the glottal instants which is further analyzed [17] to work for various other applications.

C. The DYPSA Algorithm

The miss and false alarms of the group delay method [4] have been the motivation for this method. The proposed DYPSA (Dynamic Programming projected Phase Slope Algorithm) [18] implements dynamic programming to determine the likely combination of GCIs from the GCI candidates. The GCIs are identified as positive-going zero crossings in phase slope function. Some more instants are also included in the phase slope function to reduce the miss rate. However they are considered as unity throughout the algorithm. GCI estimation is the minimization of the cost function using dynamic programming. Results of the algorithm are found to reduce false alarms by 1% and miss rates by 0.1% when compared with original GD method. The accuracy of DYPSA algorithm is 10% more than the group delay method. A modified version of DYPSA algorithm [19] implements the phase slope function and phase slope projection to estimate the GCI candidates.

Dynamic programming is used to select the candidates appropriately according to the defined cost function. The DYPSA algorithm [19] is further modified with reverberant speech [20]. Multichannel DYPSA uses correlation of GCI candidates from each microphone to provide highly robust GCI estimates. This method produces an

effective GCI estimates even in highly reverberant conditions. Further multiscale product (MP) and DYPSA algorithms are compared [21], to find the more appropriate GCI estimate and found that MP method is more effective than DYPSA. A novel approach to enhance DYPSA [22] is proposed to detect glottal instants in which the signal is derived from the multiscale product of the glottal volume flow derivative. A two-stage dynamic programming is performed to detect GCIs in the first stage and GOIs in the next stage. The block diagram for DYPSA is shown in Figure 3.

D. YAGA Algorithm

YAGA (Yet Another GCI Algorithm) [23] uses techniques based on multiscale product [24], group delay [25] and DYPSA algorithms[19]. First the set of candidates are detected from the speech signal. The effective candidates are selected by applying cost vector using dynamic programming to detect GCI. From the remaining candidates, GOIs are detected by considering the closed quotients (percentage of time the vocal folds are closed) related to the estimated GCI. The Figure 2 shows block diagram for the YAGA algorithm.

E. ZFF method:ZFF

(Zero Frequency Filtering) method is introduced for epoch extraction based on the excitation of vocal tract which produces impulse like excitation [29], with discontinuities at all frequencies. The characteristics at zero

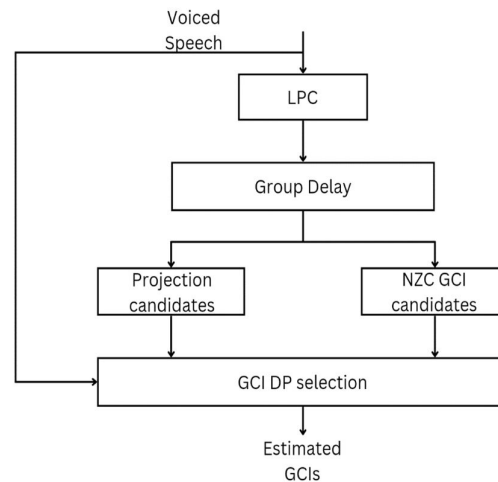


Figure 2 Block diagram for the YAGA algorithm

frequency are not affected by the resonances at higher frequencies. A speech signal is passed through two cascaded zero frequency filters to reduce the effect of all other high frequency resonances and also to have a sharper roll off. The filtered signal shows instants at positive zero crossings as epochs. Further modifying this method, zero frequency filters along with strength of excitation from the speech signal [30] are considered to estimate the epochs. Two ideal zero frequency resonators cascaded together are used to characterize the discontinuities of the impulse-like excitation. The slope of the filtered signal at the epoch reflects the strength of excitation which otherwise cannot be observed from a direct speech signal. Another method [31] is proposed that will add fixed amount of noises to the speech signal to detect high SNR instants of the glottal activity. The ZFF method is also used to detect GOIs ([32], [43]). As the ZFF method is more precise to locate GCIs, the speech signal is passed through ZFF and GCIs are obtained as ground references. The second major excitations obtained from the Hilbert Transform of the LP residual of the speech signal are identified as GOI, between two GCIs.

F. EMD: EMD

(Empirical Mode Decomposition) is a data analysis tool proposed [26], [27] to extract the frequency components of a signal. This method is implemented on speech signals [28]. EMD decomposes the speech signal into smaller set of signals, called Intrinsic Mode Functions (IMFs) [26], [27], which carry meaningful information about the signal generated. Considering the human pitch frequency range between 100Hz to 400Hz, IMFs extracted which contain the periodic information of the excitation. The first IMF represents highest frequency spectrum and the consecutive IMFs occupy lower frequency spectrum. A glottal IMF, thus represents the voiced regions of the speech signal. Further ZFF and other methods are used to identify the glottal activity.

G. GAN method:

This (Generative adversarial Network) method uses effective deep learning techniques that synthesize an EGG signal from a raw speech signal. This technique compares noisy speech signal distribution with a clean speech signal distribution and then synthesizes to get EGG-like signal. CMU-Arctic database for voiced speech signal is used [33]. The signal generated is found similar to the original dEGG signal. Sigma algorithm [11] is applied to obtain the glottal instants for the synthesized EGG-like signal.

H. Multiscale product method :

The proposed method [34] computes a three level multiscale product of speech wavelet transform at different dyadic scales in order to enhance edge detection of speech signal. The singularities or the signal peaks align along the peaks obtained from wavelet transform coefficients for the first few scales. However, choosing the scales too large will lead to misalignments as smoothing spreads the response and the singularities get separated. GCIs are identified as the greatest minimum peak obtained from the multiscale product and the smallest minimum peak situated between two GCIs is identified as a GOI. The block diagram for GD, DYPSA and MP methods is shown in Figure 3.

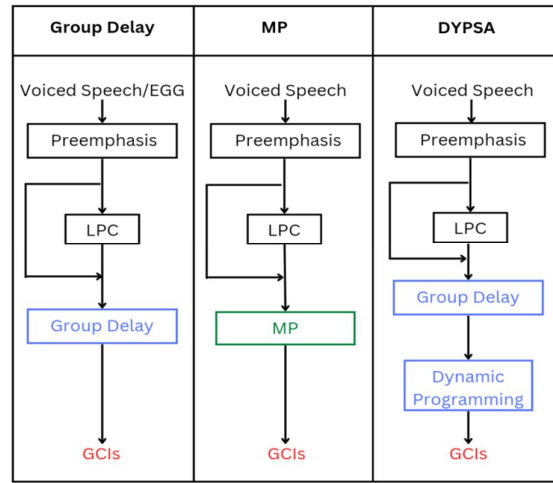


Figure 3 Block diagram for GD, DYPSA and MP methods

I. SEDREAMS method

The SEDREAMS (The Speech Event Detection using the Residual Excitation And a Mean-based Signal) method [35], implements detection of GCI by determining short intervals for the probability of GCI occurrences using mean-based signal. A mean-based signal is obtained by moving a window over the signal and finding the maxima and minima of the signal. These are the short intervals where GCIs may be found. The GCI instants are detected from the LP residual where it reaches its maximum value. Further algorithm is modified for various voice qualities [36] by applying dynamic programming to detect the GCIs.

VI. COMPARISON BASED ON THE EVALUATION OF GLOTTAL INSTANTS

Data for the researchers is provided from different data bases as APLAWD, SAM, KEELE and BDL. The methods are evaluated by the performance indices like:

1. IDR(Identification Rate): the probability of detecting atleast one instant
2. MR(Miss rate): the probability to miss an instant
3. FAR(False alarm rate): the probability of detecting more than one instant
4. IDA(Identification accuracy): Standard deviation of the identification error

The above mentioned performance indices are needed to evaluate a method or an algorithm. However, the databases and the data for various methods applied is different, hence we try to comprehend it from different papers in Table 2 and Table 3. In [19], the linear prediction coding residual [LPCR], Frobenius Norm method (FN), Group delay (GD) and Dynamic Programming Projected Phase-Slope Algorithm (DYPSA) are compared for 5 female and 5 male talkers in APLAWD and SAM databases . The small values of IDA indicate high

accuracy of detecting the instants. Hence, the GD method is found to be the best method out of LPCR and FN methods and is used as a principle GCI generator in DYPSA algorithm and the process is repeated. The performance of DYPSA over GD is found to be more accurate with identification rate of 95.7% (APLAWD) and 93.1% (SAM). Due to dynamic programming in DYPSA, rate of false alarm is also reduced. FN has a low miss rate but extremely high false alarms.

APLAWD database is used in [18] to evaluate LPCR, FN, GD and DYPSA algorithms. The enhancement of the GD algorithm by using dynamic programming reduces the miss rate and false alarm to a very great extent (0.1 and 9.9 approximately). Accuracy of the DYPSA method is found to be 90% whereas the GD is calculated as 70-80%. KEELE database is used to compare the results of [21] Multiscale Product (MP) and DYPSA for the detection of GCI. MP has a good detection rate of GCI of about 94.88%, whereas DYPSA has a detection rate of 63%. Further, failure rate of DYPSA is 31.55% where as MP is having a failure rate of 2.16%. This comparative study proves that MP method is more accurate than DYPSA algorithm. [11] Compares SIGMA, HQTx and TXGEN for detection of GCI and SIGMA, TXGEN for detection of GOI on APLAWD and SAM database. HQTx has false alarms and TXGEN has more miss errors than SIGMA algorithm. SIGMA is found to be best among the other methods.

The estimated glottal instants from the EGG signals using SIGMA algorithm and those from speech signals using DYPSA and YAGA of the APLAWD and SAM databases are compared [23]. The EGG based SIGMA algorithm performs best with high detection rate for GCI. On the other hand YAGA (99.31% in APLAWD and 98.8% in SAM databases) performs best than DYPSA (96.39% in APLAWD and 95.41% in SAM databases) in speech signal as DYPSA is dependent on the LP residual. The miss and false rates are also found to be very less for SIGMA and YAGA than DYPSA. The identification accuracy of DYPSA is found best among the SIGMA and YAGA for the detection of GOI. The group delay method followed by dynamic programming and the MP method are compared for efficiency in the GCI detection in [42] on the KEELE university and BDL databases. MP ensures the best method with identification rate of 97.8% in KEELE and 98.29% in BDL databases over the group delay method because of the high miss rate of about 15.74 in KEELE and 2.12 in BDL.

GOI estimates from EGG signal need to be extracted based on the GCIs detected. In [43], the performance of the glottal activity for SIGMA, ZFF methods on CMU-Arctic data base for EGG signals is performed. The identification rate of ZFF method is greater than the SIGMA method. Another comparison of SEDREAMS, DYPSA and ZFF based on the speech signals is performed on the CMU-Arctic database. The DYPSA algorithm for both GCI and GOI has an average identification rate as 97.88% whereas SEDREAMS has 98.69% and ZFF has 98.72%. The ZFF method has a better performance than the other two.

VII. APPLICATIONS OF GLOTTAL ACTIVITY METHODS

The accurate detection of GCI and GOI improves the performance of several speech related applications. The glottal activity detection also analyses the various types of disorders in vocal folds, which effect pitch, strength of voice, range of voice etc. Identification of GCIs in the speech signal is used in a many applications such as pitch tracking [44], prosodic speech ([45]-[46]), pathological speech [47], Speech Synthesis [50], Phonetic Factors [48], Emotional Expression [49], etc. The various methods and their possible applications are mapped in Table IV.

VIII. CONCLUSIONS

Various techniques to detect glottal activity based on EGG and speech signal are thoroughly reviewed in this paper. On comparing these techniques it is found that SEDREAMS and YAGA that process speech outperform all other techniques with respect to the identification rate on many databases. DYPSA and ZFR remain the next best methods. Considering MR and FAR, the performance of SIGMA and YAGA methods is superior to others. In terms of accuracy, YAGA leads all the algorithms. ZFR is less sensitive to reverberation and hence performs best in many of the databases. However, it is observed that very few researchers have identified the glottal activity based on the EGG signal.

REFERENCES

- [1] 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 <https://doi.org/10.1044/jsr.3103.338>
- [2] 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 [https://doi.org/10.1016/S0892-1997\(05\)80250-4](https://doi.org/10.1016/S0892-1997(05)80250-4)

TABLE II. COMPARATIVE TABLES FOR GCI

Ref. No	Database	Methods	IDR	MR	FAR	IDA
[19]	SAM	LPCR	42.3	50.5	7.2	1.47
		FN	58.7	1.4	40	0.59
		GD	82.6	4.8	12.6	0.55
		DYPSA	93.1	4	3	0.65
	APLAWD	LPCR	40.2	53.1	6.7	1.38
		FN	59.5		40.2	0.62
		GD	81.7	2.3	16	0.52
		DYPSA	95.7	1.6	2.7	0.71
[18]	APLAWD	LPCR		4.2	35.45	44.8
		FN		2.35	23.35	60.7
		GD		2.65	9.9	75.1
		DYPSA		0.1	0.85	87.7
[21]	KEELE	MPM	94.88	1.14	1.52	
		DYPSA	62.66	26.9	4.52	
[22]	APLAWD	DYPSA	96.41	1.33	2.25	0.58
[11]	APLAWD	HQTx	96.47	1.09	2.44	0.91
		TXGEN	94.78	3.47	1.76	0.93
		SIGMa-dG	99.49	0.25	0.26	0.99
		SIGMA	99.59	0.26	0.14	0.99
	SAM	HQTx	95.68	0.27	4.05	0.81
		TXGEN	90.22	9.75	0.03	0.9
		SIGMa-dG	99.27	0.34	0.39	0.98
		SIGMA	99.35	0.5	0.14	0.99
[23]	APLAWD	SIGMA	100	0	0	0.12
		DYPSA	96.39	1.54	2.07	0.65
		YAGA	99.31	0.18	0.51	0.39
	SAM	SIGMA	100	0	0	0.08
		DYPSA	95.41	2.08	2.5	0.4
[42]	KEELE	YAGA	98.8	0.43	0.77	0.31
		GD+DP	81.3	15.74	2.98	0.55
	BDL	MP	97.8	1.83	0.38	0.3
		GD+DP	95.54	2.12	2.34	0.42
		MP	98.29	1.62	0.11	0.21
[43]	BDL	SIGMA	94.82	3.8	1.38	0.35
		ZFF	100	0	0	0.25
	JMK	SIGMA	96.48	1.73	1.79	0.35
		ZFF	100	0	0	0.28
	SLT	SIGMA	97.97	0.93	1.1	0.31
		ZFF	100	0	0	0.21
	Average	SIGMA	96.42	2.15	1.43	0.34
		ZFF	100	0	0	0.25

TABLE III. COMPARATIVE TABLES FOR GOI

Ref. No	Database	Methods	IDR	MR	FAR	IDA
[22]	APLAWD	DYPSA	95	1.9	3.09	1.09
[11]	APLAWD	TXGEn	94.38	3.55	1.82	0.93
		SIGMA-dG	99.38	0.3	0.32	0.98
		SIGMA	99.47	0.32	0.21	0.99
	SAM	TXGEn	90.44	9.44	0.11	0.9
		SIGMA-dG	98.95	0.34	0.71	0.98
		SIGMA	99.23	0.38	0.39	0.98
[23]	APLAWD	SIGMA				0.7
		DYPSA				0.53
		YAGA				0.63
	SAM	SIGMA				0.53
		DYPSA				0.39
[43]	BDL	YAGA				0.54
		SIGMA	95.68	2.68	1.34	0.52
		ZFF	99.42	0.33	0.25	0.71
	JMK	SIGMA	96.49	1.65	1.86	0.71
		ZFF	99.59	0.24	0.17	0.67
	SLT	SIGMA	97.77	0.93	1.3	0.58
		ZFF	99.96	0.02	0.02	0.51

	Average	SIGMA	96.75	1.75	1.5	0.6
		ZFF	99.66	0.19	0.15	0.63
	BDL	DYPSA	97.25	0.35	2.4	0.93
		SEDREAMS	98.03	0.39	1.58	0.77
		ZFF	98.35	0.89	0.76	0.74
	JMK	DYPSA	98.77	0.27	0.96	0.69
		SEDREAMS	99.37	0.29	0.34	0.79
		ZFF	98.4	1.37	0.23	0.91
	SLT	DYPSA	97.88	0.5	1.62	0.55
		SEDREAMS	98.63	0.91	0.46	0.7
		ZFF	99.2	0.24	0.56	0.71
	Average	DYPSA	97.88	0.48	1.68	0.72
		SEDREAMS	98.69	0.56	0.78	0.75
		ZFF	98.72	0.89	0.48	0.78

TABLE IV APPLICATIONS OF VARIOUS GLOTTAL ACTIVITY DETECTION METHODS

Methods	Applications					
	Pitch Tracking	Prosodic Speech Modification	Pathological Speech	Speech Synthesis	Phonetic Factors	Emotional Expression
Group delay	X	X		X	X	
Multiscale Product	X		X			
SIGMA	X					
LOMA	X			X		
DYPSA	X	X			X	X
EMD	X					
ZFF	X	X	X	X		X
SEDREAMS	X	X		X		

- [3] M. A. Huckvale, "Speech Filing System: Tools for Speech," Tech. Rep. Univ. College London, London, U.K., 2004 [Online]
- [4] R. Smits and B. Yegnanarayana, "Determination of instants of significant excitation in speech using group delay function," IEEE Trans. Speech Audio Process., vol. 3, no. 5, pp. 325–333, Sep. 1995 doi: 10.1109/89.466662
- [5] B. Yegnanarayana and R. Smits, "A robust method for determining instants of major excitations in voiced speech," in Proc. ICASSP, Detroit, MI, 1995 doi: 10.1109/ICASSP.1995.479809
- [6] P. S. Murthy and B. Yegnanarayana, "Robustness of group-delay-based method for extraction of significant instants of excitation from speech signals," IEEE Trans. Speech Audio Process., vol. 7, pp. 609–619, Nov. 1999 doi: 10.1109/89.799686
- [7] Mike Brookes, Member, IEEE, Patrick A. Naylor, Member, IEEE A Quantitative Assessment of Group Delay Methods for Identifying Glottal Closures in Voiced Speech, 2006 doi: 10.1109/TSA.2005.857810
- [8] A. Bouzid, and N. Ellouze, Local Regularity Analysis at Glottal Opening and Closing Instants in Electroglottogram Signal Using Wavelet Transform Modulus Maxima, in Proc. Eurospeech 2003, Geneve, pp. 2837–2840, 2003
- [9] Electroglottographic Measures Based on GCI and GOI Detection Using Multiscale Product Aicha Bouzid, Nouredine Ellouze, 2008
- [10] A. Rosenfeld, A Non-Linear Edge Detection, Proc. IEEE, Vol. 58, pp. 814–816, 1970 doi: 10.1109/PROC.1970.7756
- [11] Mark R. P. Thomas, and Patrick A. Naylor The SIGMA Algorithm: A Glottal Activity Detector for Electroglottographic Signals doi: 10.1109/TASL.2009.2022430
- [12] Pramod B. Bachhav, Hemant A. Patil and Tanvina B. Patel A novel filtering based approach for epoch extraction, 2015 doi: 10.1109/ICASSP.2015.7178879
- [13] S.B. Sunil Kumar, Tanumay Mandal, K. Sreenivasa Rao, Robust glottal activity detection using the phase of an electroglottographic signal, Biomedical Signal Processing and Control, Volume 36, 2017, Pages 27–38, ISSN 1746-8094, <https://doi.org/10.1016/j.bspc.2017.03.007>.
- [14] T. Mandai and K. S. Rao, "Robust Detection of Glottal Activity Using Unwrapped Phase Electroglottographic Signal," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5584–9, doi: 10.1109/ICASSP.2018.8461565.
- [15] Vu Ngoc Tuan & Christophe d'Alessandro Robust glottal closure detection using the wavelet transform., 6th European Conference on Speech Communication and Technology (EUROSPEECH'99)
- [16] N. Sturmel, C. d'Alessandro and F. Rigaud, "Glottal closure instant detection using Lines of Maximum Amplitudes (LOMA) of the wavelet transform," 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, pp. 4517–4520, doi: 10.1109/ICASSP.2009.4960634

- [17] Christophe D'alessandro* And Nicolas Sturm Glottal closure instant and voice source analysis using time-scale lines of maximum amplitude, 2011 <https://doi.org/10.1007/s12046-011-0040-6>
- [18] Anastasis Kounoudes, Patrick A. Naylor, Mike Brookes The DYPSA algorithm for estimation of glottal closure instants in voiced speech, 2002 doi: 10.1109/ICASSP.2002.5743726
- [19] Patrick A. Naylor, Member, IEEE, Anastasis Kounoudes, Member, IEEE, Jon Gudnason, Member, IEEE, and Mike Brookes Estimation of Glottal Closure Instants in Voiced Speech Using the DYPSA Algorithm, 2007 doi: 10.1109/TASL.2006.876878
- [20] Mark R. P. Thomas, Nikolay D. Gaubitch and Patrick A. Naylor Multichannel DYPSA for estimation of glottal closure instants in reverberant speech, 2007
- [21] WafaSaïdi, Aïcha Bouzid, Nouredine Ellouze Evaluation of Multi-scale Product Method and DYPSA Algorithm for Glottal Closure Instant Detection, 2008 doi: 10.1109/ICTTA.2008.4530031
- [22] Mark R. P. Thomas, Jon Gudnason and Patrick A. Naylor Detection of glottal closing and opening instants using an improved DYPSA framework, 2009
- [23] Mark R. P. Thomas, Member, IEEE, Jon Gudnason, Member, IEEE, and Patrick A. Naylor Estimation of Glottal Closing and Opening Instants in Voiced Speech Using the YAGA Algorithm, 2012 doi: 10.1109/TASL.2011.2157684
- [24] A. Bouzid and N. Ellouze, "Open quotient measurements based on multiscale product of speech signal wavelet transform," *Res. Lett. Signal Process.*, 2007 <https://doi.org/10.1155/2007/62521>
- [25] M. Brookes, P. A. Naylor, and J. Gudnason, "A quantitative assessment of group delay methods for identifying glottal closures in voiced speech," *IEEE Trans. Speech Audio Process.*, vol. 14, no. 2, pp. 456–466, Mar. 2006 doi: 10.1109/TSA.2005.857810
- [26] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.- C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998 <https://doi.org/10.1098/rspa.1998.0193>
- [27] N. E. Huang and S. S. Shen, *Hilbert-Huang transform and its applications*. World Scientific, 2005, vol. 5 https://doi.org/10.1142/9789812703347_0009
- [28] Rajib Sharma and S. R. Mahadeva Prasanna Characterizing Glottal Activity from Speech using Empirical Mode Decomposition, 2015 doi: 10.1109/NCC.2015.7084821.
- [29] K. Sri Rama Murty and B. Yegnanarayana Epoch Extraction From Speech Signals, 2008, doi: 10.1109/TASL.2008.2004526
- [30] Characterization of Glottal Activity From Speech Signals K. Sri Rama Murty, B. Yegnanarayana, Senior Member, IEEE, and M. Anand Joseph, 2009 doi: 10.1109/LSP.2009.2016829.
- [31] Voiced/Nonvoiced Detection Based on Robustness of Voiced Epochs N. Dhananjaya and B. Yegnanarayana, 2010 doi: 10.1109/LSP.2009.2038507
- [32] Ramesh K.1, S. R. M. Prasanna1 and D. Govind Detection of Glottal Opening Instants using Hilbert Envelope, 2013
- [33] K.T.Deepak, Pavitra Kulkarni, U Mudenagudietal Glottal Instants extraction from speech signal using Generative Adversarial network.[2019] doi: 10.1109/ICASSP.2019.8683298
- [34] Aicha Bouzid and Nouredine Ellouze Glottal opening instant detection from speech signal, 2019
- [35] T. Drugman, M. Thomas, J. Gudnason, P. Naylor and T. Dutoit, "Detection of glottal closure instants from speech signals: A quantitative review", *IEEE Trans. on Audio Speech and Lang. Process.*, vol. 20, no. 3, pp. 994–1006, 2012. doi: 10.1109/TASL.2011.2170835
- [36] J. Kane and C. Gobl, "Evaluation of glottal closure instant detection in a range of voice qualities", *Speech Communication*, vol. 55, no. 2, pp. 295–314, 2013 <https://doi.org/10.1016/j.specom.2012.08.011>
- [37] V. Khanagha, K. Daoudi and H. Yahia, "Detection of glottal closure instants based on the microcanonical multiscale formalism", *IEEE/ACM Trans. on Audio Speech and Lang. Process.*, vol. 22, no. 12, pp. 1941–1950, Dec 2014. doi: 10.1109/TASLP.2014.2352451
- [38] Arturo Camacho and John G. Harris, "A sawtooth waveform inspired pitch estimator for speech and music", *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1638–1652, 2008. <https://doi.org/10.1121/1.2951592>
- [39] T. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-27, no. 4, pp. 309–319, Aug. 1979. doi: 10.1109/TASSP.1979.1163267
- [40] Y. M. Cheng and D. O'Shaughnessy, "Automatic and reliable estimation of glottal closure instant and period," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1805–1815, Dec. 1989. doi: 10.1109/29.45529
- [41] 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 Process. Lett.*, vol. 14, no. 10, pp. 762–765, Oct. 2007. doi: 10.1109/LSP.2007.896454
- [42] Group Delay Function Followed by Dynamic Programming Versus Multiscale-Product for Glottal Closure Instant Detection GhayaSmidi and Aicha Bouzid, 2019, https://doi.org/10.1007/978-981-15-0637-6_19
- [43] Ramesh, K., Prasanna, S.R.M. Glottal opening instants detection using zero frequency resonator. *Int J Speech Technol* 20, 127–141 (2017). <https://doi.org/10.1007/s10772-016-9383-z>
- [44] E. Loweimi J. Barker T. Hain On the usefulness of the speech phase spectrum for pitch extraction, January 2018, International Speech Communication Association' DOI:10.21437/Interspeech.2018-1062

- [45] Govind, D., Joy, T.T. Improving the Flexibility of Dynamic Prosody Modification Using Instants of Significant Excitation. *Circuits Syst Signal Process* 35, 2518–2543 (2016). <https://doi.org/10.1007/s00034-015-0159-5>
- [46] Vekkot, S., Gupta, D. Prosodic transformation in vocal emotion conversion for multi-lingual scenarios: a pilot study. *Int J Speech Technol* 22, 533–549 (2019). <https://doi.org/10.1007/s10772-019-09626-5>
- [47] , Xuehui Zhang, Weiping Hu Pathological Voice Classification Based on Wavelet Packet Multiscale Analysis, *ACAI 2018: Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, December 2018 Article No.: 13Pages 1–6<https://doi.org/10.1145/3302425.3302450>
- [48] Vainio, M. (2014). *Phonetics and Machine Learning: Hierarchical Modelling of Prosody in Statistical Speech Synthesis*. *Lecture Notes in Computer Science*, 37–54. doi:10.1007/978-3-319-11397-5_3
- [49] Vekkot, S., Tripathi, S. (2018). Significance of Glottal Closure Instants Detection Algorithms in Vocal Emotion Conversion. In: Balas, V., Jain, L., Balas, M. (eds) *Soft Computing Applications*. *SOFA 2016. Advances in Intelligent Systems and Computing*, vol 633. Springer, Cham. https://doi.org/10.1007/978-3-319-62521-8_40
- [50] Lal, G.J., Gopalakrishnan, E.A. & Govind, D. Epoch Estimation from Emotional Speech Signals Using Variational Mode Decomposition. *Circuits Syst Signal Process* 37, 3245–3274 (2018). <https://doi.org/10.1007/s00034-018-0804-x>