COMPARATIVE STUDY OF DIFFERENT EPOCH EXTRACTION METHODS FOR SPEECHASSOCIATED WITH VOICE DISORDERS

by

Purva Barche, Krishna Gurugubelli, Anil Kumar Vuppala

Report No: IIIT/TR/2021/-1

Centre for Language Technologies Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
June 2021

COMPARATIVE STUDY OF DIFFERENT EPOCH EXTRACTION METHODS FOR SPEECH ASSOCIATED WITH VOICE DISORDERS

Purva Barche, Krishna Gurugubelli, and Anil Kumar Vuppala

Speech Processing Laboratory, LTRC, KCIS
International Institute of Information Technology, Hyderabad, India.
Email: {purva.sharma, krishna.gurugubelli}@research.iiit.ac.in, and anil.vuppala@iiit.ac.in

ABSTRACT

Accurate detection of epoch locations is important in extracting the features from the speech signal for automatic detection and assessment of voice disorders. Therefore, this study aimed to compare the various algorithms for detecting epoch locations from the speech associated with voice disorders. In this regard, nine state-of-the-art epoch extraction algorithms were considered, and their performance for different categories of voice disorders was evaluated on the SVD dataset. Experimental results indicate that most of the epoch extraction methods showed better performance for healthy speech; however, their performance was degraded for speech associated with voice disorders. Furthermore, the performance of epoch extraction methods was degraded for the speech of structural and neurogenic disorders compared to the speech of psychogenic and functional disorders. Among the different epoch extraction algorithms, zero phase-zero frequency filtering showed the best performance in terms identification rate (90.37%) and identification accuracy (0.34ms), for speech associated with voice disorders.

***Index Terms*—** Epoch locations, Excitation source, Glottal closure instants, Voice disorders.

1. INTRODUCTION

Speech production is a complex process that involves respiration, phonation, resonance and articulation. Phonation is a process in which vocal folds produce quasi-periodic vibration; this process is responsible for sound to be audible [1]. Voice disorders may occur due to poor respiratory system, or incomplete glottal closure or extra lesion on vocal fold or irregularity in the vibration of vocal fold or abnormal vocal fold closure or weakness in muscles which are responsible for voicing or it may be due to psychological reason [2]. Voice disorders affect the phonation, which is manifested as irregularities in vocal fold vibration during speech production [2, 3]. The information related to voice disorders is mainly captured in excitation source [4], and the knowledge of epoch locations is useful in extracting the excitation source from speech [5]. Epoch locations also referred to as glottal closure instants (GCIs), are the instants at which significant excitation of the vocal tract system happens during speech phonation [6]. The knowledge of epoch locations is useful in glottal source analysis [5], efficient estimation of vocal tract system information [7], and pathological speech analysis [8, 9, 10]. Excitation source features like glottal parameters, fundamental frequency (F0) parameters, and jitter measurements are useful in the detection and assessment of voice disorders [11, 12]. Computation of these features involves the detection of epoch locations from speech. Hence, accurate estimation of epoch locations is important for computing the features for the automatic detection and assessment of voice disorders.

The studies in [13, 14], show that the performance of state-of-the-art epoch extraction methods is efficient in clean speech conditions. Efficacy of epoch extraction methods has been studied for telephonic quality speech [15, 16, 17], emotional speech [18, 19], and the degraded speech obtained by corrupting the clean speech with additive noise and reverberations [20, 21]. In general, the performance of these methods has been evaluated using speech utterances produced by healthy (controlled) speakers. On the other hand, the subjects suffering from voice disorders will not be able to produce normal or model phonation [3]. Hence, the performance of existing epoch extraction methods may vary in the processing of speech associated with voice disorders due to the variations in the glottal source characteristics such as roughness, breathiness, hoarseness, abnormality in pitch and strained quality [2, 4]. In literature, the performance of epoch extraction methods was not studied for the speech associated with voice disorders. Hence, the present study aims to compare the performance of various state-of-the-art algorithms for extracting epoch locations from speech associated with voice disorders. Moreover, the performance of a GCI detection method may vary depending on the type of voice disorder because each voice disorder can affect the phonation process in a different way. Hence, this study also intended to investigate the performance of the epoch extraction methods for different voice disorders by using Saarbruecken voice disorder database [22].

Rest of the paper is organised as follows: Section 2 describes the categories of voice disorders from a clinical perspective. Section 3 presents the experimental setup, which includes the database, followed by the epoch extraction methods and the evaluation metrics. Results and discussions of different epoch extraction methods are presented in Section 4. Finally, the summary and conclusions of the study are described in Section 5.

2. CATEGORIZATION OF VOICE DISORDERS: FROM A CLINICAL PERSPECTIVE

This study intended to study the performance of epoch extraction algorithms for the speech of different voice disorders. There are many voice disorders which affect the phonation process while producing the speech. Hence this section describes the classification of voice disorders from a clinical perspective. The typical symptoms of a voice disorder include degradation of an individual's voice quality, reduction in loudness, loss of voice, and increase in vocal effort [2]. According to *American speech language hearing association* (ASHA), voice disorders are categorized into organic and non-organic [23]. Organic voice disorders are mainly due to anatomic abnormality in the larynx or due to the neurological damage [3]. Organic voice disorders are categorized into two sub-types: Structural and Neurogenic.

- Structural voice disorders (Polyp, Nodules, Leukoplakia, Laryngitis) are mainly due to anatomic abnormalities (like growth of the lesion, swelling of vocal cords) in the larynx [2].
- Neurogenic voice disorders (Spasmodic dysphonia and recurrent laryngeal nerve palsy) are caused due to the damage or malfunction in central or peripheral nervous system [24]. As the nervous system interacts with the larynx, it affects the functioning of the vocal mechanism.

On the other hand, non-organic disorders are either due to strain in the muscle or/and psychological reason. The non-organic disorders categorized into two types: Functional and Psychogenic.

- Functional voice disorders (commonly known as Muscle Tension Dysphonia) are characterized by excessive laryngeal activity, tension, reduced vocal capacity, and impaired voice without any organic abnormality [25].
- In psychogenic voice disorders subject will lose control over the initiation and maintenance of phonation during speech production due to disturbed psychological process like anxiety, depression, conversion reaction, or personality disorder [26, 27].

Comparative analysis of state-of-the-art GCI detection algorithms was studied on four broad categories of voice disorders, namely, structural, neurogenic, functional, and psychogenic.

3. EXPERIMENT SETUP

This section briefly describes the database, and the state-of-the-art epoch extraction methods considered for the current study.

3.1. Database

In the Saarbruecken voice disorder (SVD) database [22], for each of the speech recordings, simultaneous EGG signals are available to obtain the ground truth epoch locations. Therefore, this study used the SVD database to evaluate the performance of epoch extraction algorithms. This is a publicly available database, can be downloaded from the site `http://www.stimmdatenbank.coli.uni-saarland.de/`. The present study considered the speech recordings from 687 healthy subjects and 679 subjects with different voice disorders from the SVD database. Each recording includes vowels /a/, /i/ and /u/ produced at a normal, low, and high pitch and also with rising-falling pitch. Also, each recording consists of a German sentence “Guten Morgen, wie geht es Ihnen?” (“Good morning, how are you?”). The SVD database was recorded at a sampling frequency of 50 kHz. In this study, all recordings were down-sampled to 8000 Hz. Additionally, the speech recordings correspond to 679 subjects with different voice disorders were categorized into four sub-classes, namely, Structural, Neurogenic, Functional, and Psychogenic (as discussed in Section 2). Further details, about each of the sub-classes are provided in Table 1.

3.2. Epoch Extraction Methods

Brief description of each epoch extraction methods considered for this study is explained as follows.

Table 1. Details of the voice disorders considered from SVD database for evaluating the epoch extraction algorithms. Here, FD: Functional dysphonia, PD: Psychogenic dysphonia, RLNP: Recurrent laryngeal nerve palsy, and SD: Spasmodic dysphonia,

Voice disorder type	Disorder name	#Speakers
Structural	Laryngitis	30
	Leukoplakia	41
	Polyp	45
Neurogenic	SD	30
	RLNP	188
Functional	FD	254
Psychogenic	PD	91

3.2.1. ZFF Method

Zero frequency filtering (ZFF) algorithm is based on the fact that discontinuity which occurs at each glottal closure instant reflects across all the frequency including the zero frequency [6]. Hence, in ZFF method, the speech signal is passed through a zero frequency resonator to attenuates the effect of higher frequency resonances due to the vocal tract system, which in turn highlights the excitation source characteristics. The resonator used in ZFF approach is a narrow band filter of order four with poles located at the unit circle. The output of the ZFF filter has polynomial growth/decay. The trend in the filter’s output is removed by subtracting it from the local mean. The mean subtracted signal is referred to as zero frequency filtered signal. And, the positive to negative zero crossings of the zero frequency filtered signal are referred to as epochs.

3.2.2. ZP-ZFF Method

On the other hand, *zero phase-zero frequency filtering* (ZP-ZFF) is an alternative to ZFF approach [28]. The zero frequency resonator used in this method is non-causal, stable and IIR filter. The ZP-ZFF provides the stable implementation of ZFF, to estimate epochs from speech. More details of this implementation can be seen in [28].

3.2.3. DYPSA Method

The *dynamic programming phase slope algorithm* (DYPSA) uses the excitation source information for the detection of GCIs [29]. Epoch detection procedure was performed in three steps: group delay function, phase slope projection and dynamic programming. First, the group delay function is calculated from the LP residual signal. Then GCIs are estimated from negative zero crossings of group delay function. Then phase slope projection is used to estimate the GCIs which are missed out by group delay function. From these two parts, most of the true GCIs are selected, but this procedure also results in false GCIs. Hence, to select the true GCIs, dynamic programming algorithm is used, which minimizes the various cost functions.

3.2.4. YAGA Method

The *yet another GCI algorithm* (YAGA) uses wavelet analysis, group delay analysis and M-best dynamic programming to estimate the epoch locations from speech [30]. In this regard, the multi-scale product of the stationary wavelet transform is utilized to highlight the discontinuities in the voice source signal. Then, the GCIs are estimated through iterative adaptive inverse filtering approach by using a voice source signal. Finally, M-best dynamic programming is performed to minimize false epoch locations.

3.2.5. SEDREAMS Method

In the *speech event detection using the residual excitation and a mean-based signal* (SEDREAMS) method, epoch locations are estimated in two steps [13]. First, a mean based signal is estimated by computing the moving average of speech signal using a Blackman window. Then the mean based signal is used to compute short intervals, where GCIs are expected to occur. Finally, the short intervals are refined to find out the exact GCI locations. In this regard, a peak picking algorithm is applied on the LP residual, assuming the largest discontinuity within an interval corresponds to GCIs.

3.2.6. SE-VQ Method

To handle the different phonation type, *SEDREAMS-voice quality* (SE-VQ) algorithm was proposed, which is a modified form of SEDREAMS algorithm [31]. In this method, two extra steps are introduced as compare to basic SEDREAMS algorithm they are: dynamic programming and post-processing. Dynamic programming is applied to select the optimal GCI locations based on the strength of peaks in LP residual and transition cost (i.e. transition from one GCI to another GCI). Further, post-processing is applied to minimize the false positive GCIs location and to preserves the true positive GCIs. In the SEDREAMS only one peak which is the highest peak from LP residual is chosen, while in the SE-VQ, several LP residual peaks are selected in order to handle the voice quality like breathy and harsh where there are no prominent peaks.

3.2.7. GEFBA Method

Glottal closure/opening instant estimation using forward-backward algorithm (GEFBA) is based on source signal obtained by linear prediction based inverse filtering [20]. This algorithm is performed in two phases. In the first phase, the glottal flow derivative is derived from inverse filtering based on LP analysis. Finally, in the second phase of GEFBA algorithm, a forward and move backward algorithm is performed on each voiced frame to estimate GCIs.

3.2.8. CWT-GCI Method

Continuous wavelet transform-glottal closure instant (CWT-GCI) algorithm is based on the principle that CWT is a suitable method for determining the sharp transition from the signal [16]. In this method, to compute GCIs CWT coefficients are calculated from the analytic signal instead of from the speech signal. From these coefficients, the average absolute signal is obtained, and this signal is convoluted with a Gaussian filter to highlights the peaks. The convoluted output is referred to as evidence to estimate the epoch locations. Spurious peaks are removed from the evidence signal by considering that time difference between the two consecutive peaks is not less than 2 ms. After removing the spurious peaks, positive peaks obtained from epoch evidence signal are referred to as epoch locations.

3.2.9. SPF Method

This algorithm of epoch extraction is based on the estimation of time-frequency representation obtained from single pole filter (SPF) [15]. Single pole filter is a narrow band IIR filter, with pole located inside the unit circle. In this approach, first, the speech signal is passed through the bank of single-pole filters, which gives better time-frequency representation of the speech signal. From this time-frequency representation, time marginal is derived. Further, the time marginal is smoothed using a Gaussian window of 8 ms. Finally, positive crossings obtained from the smoothed time marginal, which are referred to as epoch locations.

For reproducing the results of the current study, MATLAB implementations of all epoch excitation algorithms considered in this study are provided in the following link `https://github.com/gurugubelllik/Epoch-extraction-methods.git`.

3.3. Evaluation Metrics

The performance of the epoch extraction algorithms is evaluated by using four standard measures which are defined as:

- Identification rate (IDR in %) is defined percentage of the larynx cycles in which exactly one GCI is detected.
- Miss rate (MR in %) is defined as the percentage of the larynx cycle in which no GCI is detected.
- False alarm rate (FAR in %) is defined as a percentage of the larynx cycle in which more than one GCI is detected.
- Identification accuracy (IDA in ms) is defined as the standard deviation of the time between reference and detected GCI in the larynx cycle only for which exactly one GCI has been detected.

More details about these measures can be found in [30].

4. RESULTS AND DISCUSSION

In this section, we compared the performance of nine state-of-the-art epoch extraction methods for the speech of healthy subjects and the speech of subjects with various voice disorders, using the SVD database, which provides simultaneous EGG recordings. The performance of each method is evaluated in terms of IDR, MR, FAR and IDA. Evaluation measures of algorithms extract epoch locations from healthy speech and speech associated with voice disorders were calculated and reported in Table 2. In addition, the performance of the epoch extraction algorithms was studied for each of the four broad categories of voice disorders (structural, neurogenic, functional, and psychogenic), and the evaluation measures were reported in Table 3.

From the results presented in Table 2, it is evident that most of the epoch extraction methods (except SE-VQ, CWT-GCI, SPF, and GEFBA) work well for the healthy scenario, in which speech is produced under model phonation. However, all epoch extraction methods show significant degradation in their performance for speech associated with voice disorders compared to healthy speech. Compared to the healthy scenario, in the voice disorder scenario, all the epoch extraction algorithms shown approximately 5 to 8% absolute reduction in IDR and (0.05 to 0.15) ms absolute increase in IDA. Among all epoch extraction methods, SEDREAMS and ZP-ZFF methods performed better in both healthy and voice disorder scenarios, in terms of IDR, FAR, and IDA. In the healthy scenario, SEDREAMS method shows the best performance in terms of IDR of 97.69%, whereas, ZP-ZFF method shown to be second best with an IDR of 97.63%. In the voice disorder scenario, the ZP-ZFF method showed the best performance in terms of IDR of 90.37% and IDA of 0.34 ms, while the ZFF and SEDREAMS methods showed IDR of 89.96% each one, which is almost equivalent to the IDR of ZP-ZFF. On the other hand, the DYPSA and YAGA methods showed comparable results in terms of IDR.

Table 2. Performance evaluation of different epoch extraction methods for speech of healthy speakers and speech of speakers with voice disorder on SVD dataset. IDR–Identification rate, MR–Miss rate, FAR–False Alarm Rate, IDA–Identification Accuracy

Class	Method	IDR (%)	MR (%)	FAR (%)	IDA (ms)
Healthy	ZP-ZFF	97.63	1.16	**1.21**	0.26
	ZFF	96.94	**0.75**	2.31	0.42
	DYPSA	95.45	1.42	3.13	**0.23**
	YAGA	96.22	1.03	2.75	0.66
	SEDREAMS	**97.69**	0.87	1.44	0.28
	SE-VQ	78.36	16.12	5.52	0.85
	CWT-GCI	92.01	6.35	1.65	0.45
	SPF	87.19	10.47	2.34	0.43
	GEFBA	72.77	22.09	5.14	0.54
Voice disorders	ZP-ZFF	**90.37**	4.03	5.6	**0.34**
	ZFF	89.96	3.79	6.25	0.46
	DYPSA	88.06	4.57	7.37	0.36
	YAGA	88.1	**3.62**	8.28	0.68
	SEDREAMS	89.96	4.44	5.59	0.39
	SE-VQ	74.01	19.05	6.93	0.91
	CWT-GCI	85.77	9.64	**4.59**	0.56
	SPF	81.27	13.79	4.93	0.59
	GEFBA	64.96	27.02	8.01	0.58

From the results reported in Table 3, it can be understood that among all the categories of voice disorders for the structural and neurogenic categories, the performance of all epoch extraction algorithms was very poor in terms of identification rate. Compared to the healthy scenario, for the structural, neurogenic, functional, and psychogenic voice disorder scenarios the epoch extraction algorithms showed an absolute reduction in IDR of approximately 10%, 15%, 3%, and 5%, respectively. The IDA refers to standard deviation of error, and therefore it should be lower for better performance of an epoch extraction method [6]. The IDA of the epoch extraction methods in neurogenic and structural voice disorder scenarios was increased approximately by 20 ms. More interestingly, the performance of epoch extraction methods degraded more for organic voice disorders (structural and neurogenic) than for non-organic voice disorders (functional and psychogenic). The results of this study indicate that existing epoch extraction methods need to be improved for the accurate detection of epoch locations from the speech in the context of voice disorders.

5. SUMMARY AND CONCLUSION

Due to the vital importance of epoch locations in obtaining the features for the automatic detection and assessment of voice disorders, this study aimed to compare the performance of epoch extraction algorithms for speech associated with voice disorders. In this regard, nine state-of-the-art epoch extraction algorithms were considered for the analysis and their performance was evaluated using the Saarbruecken voice disorder database. Based on the results of the present study, it is understood that all epoch extraction algorithms are shown degradation in their performance for speech associated with voice disorders compared to healthy speech. Most importantly, the existing algorithms are shown poor performance in structural and neurogenic voice disorder scenarios. Among all epoch extraction methods, ZP-ZFF method shown the best performance in the voice disorder scenario, in terms of IDR (90.37%) and IDA (0.34 ms). In our future works, we intended to explore different post-processing techniques to improve the performance of existing epoch extraction methods for structural and neurogenic voice disorder scenarios.

Table 3. Performance evaluation of different epoch extraction methods for speech associated with different types of voice disorders on SVD dataset. IDR–Identification rate, MR–Miss rate, FAR–False Alarm Rate, IDA–Identification Accuracy

Class	Method	IDR (%)	MR (%)	FAR (%)	IDA (ms)
Structural Disorders	ZP-ZFF	**87.84**	5.72	6.44	0.42
	ZFF	87.79	5.37	6.84	0.52
	DYPSA	84.53	5.86	9.61	**0.41**
	YAGA	85.48	**4.91**	9.61	0.84
	SEDREAMS	87.51	6.24	6.25	0.47
	SE-VQ	74.63	17.97	7.40	1.02
	CWT-GCI	86.10	8.30	5.59	0.62
	SPF	83.57	10.9	**5.54**	0.66
	GEFBA	70.77	20.47	8.76	0.64
Neurogenic Disorders	ZP-ZFF	**84.04**	7.04	8.92	**0.42**
	ZFF	83.33	6.90	9.77	0.55
	DYPSA	81.40	7.32	11.28	0.47
	YAGA	81.76	**6.22**	12.03	0.71
	SEDREAMS	83.32	8.24	8.44	0.49
	SE-VQ	71.02	20.88	8.10	1.00
	CWT-GCI	80.83	11.90	7.27	0.65
	SPF	77.35	15.60	**7.04**	0.69
	GEFBA	61.06	30.89	8.05	0.62
Functional Disorders	ZP-ZFF	**95.54**	1.40	3.07	**0.27**
	ZFF	95.28	**1.17**	3.56	0.40
	DYPSA	93.80	2.16	4.03	0.28
	YAGA	93.10	1.40	5.50	0.62
	SEDREAMS	95.41	1.23	3.36	0.30
	SE-VQ	76.18	17.72	6.10	0.84
	CWT-GCI	88.97	8.59	**2.44**	0.49
	SPF	82.99	13.54	3.47	0.50
	GEFBA	65.92	26.16	7.92	0.55
Psychogenic Disorders	ZP-ZFF	**94.34**	2.00	3.66	**0.27**
	ZFF	93.77	**1.67**	4.56	0.37
	DYPSA	92.49	3.06	4.44	0.29
	YAGA	92.70	1.92	5.38	0.56
	SEDREAMS	93.81	2.00	4.18	0.28
	SE-VQ	74.36	19.76	5.88	0.77
	CWT-GCI	88.28	8.86	**2.86**	0.46
	SPF	82.93	13.85	3.21	0.51
	GEFBA	64.23	28.53	7.24	0.50

6. REFERENCES

[1] Raymond D Kent, Charles Read, and Ray D Kent, *The acoustic analysis of speech*, vol. 58, Singular Publishing Group San Diego, 1992.

[2] A E Aronson, "Clinical voice disorders," *An interdisciplinary approach*, 1985.

[3] M Ed Freeman and M Ed Fawcus, *Voice disorders and their management*, Whurr Publishers, 2000.

[4] Youri Maryn, Marc De Bodt, and Nelson Roy, "The acoustic

voice quality index: toward improved treatment outcomes assessment in voice disorders," *Journal of Communication Disorders*, vol. 43, no. 3, pp. 161–174, 2010.

[5] Alku Paavo, "Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering," *Speech communication*, vol. 11, no. 2-3, pp. 109–118, 1992.

[6] K S R Murty and B Yegnanarayana, "Epoch extraction from speech signals," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 16, no. 8, pp. 1602–1613, 2008.

[7] M Airaksinen, T Raitio, B Story, and Alku Paavo, "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 22, no. 3, pp. 596–607, 2014.

[8] D G Silva, L C Oliveira, and M Andrea, "Jitter estimation algorithms for detection of pathological voices," *EURASIP Journal on Advances in Sig. Process.*, vol. 2009, pp. 1–9, 2009.

[9] P Gómez-Vilda, R Fernández-Baillo, V Rodellar-Biarge, V N Lluis, A Álvarez-Marquina, L M Mazaira-Fernández, R Martínez-Olalla, and J I Godino-Llorente, "Glottal source biometrical signature for voice pathology detection," *Speech Communication*, vol. 51, no. 9, pp. 759–781, 2009.

[10] M H Javid, Krishna Gurugubelli, and Anil Kumar Vuppala, "Single frequency filter bank based long-term average spectra for hypernasality detection and assessment in cleft lip and palate speech," in *Proc. ICASSP*. IEEE, 2020, pp. 6754–6758.

[11] John Laver, Steven Hiller, and Janet Mackenzie Beck, "Acoustic waveform perturbations and voice disorders," *Journal of Voice*, vol. 6, no. 2, pp. 115–126, 1992.

[12] Purva Barche, Krishna Gurugubelli, and Anil Kumar Vuppala, "Towards automatic assessment of voice disorders: A clinical approach," *Proc. Interspeech 2020*, pp. 2537–2541, 2020.

[13] 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, 2011.

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

[15] C M Vikram and S R M Prasanna, "Epoch extraction from telephone quality speech using single pole filter," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 25, no. 3, pp. 624–636, 2017.

[16] Y M Keerthana, M K Reddy, and K S Rao, "Cwt-based approach for epoch extraction from telephone quality speech," *IEEE Sig. Process. Lett.*, vol. 26, no. 8, pp. 1107–1111, 2019.

[17] Krishna Gurugubelli, H M Javid, KNRK Raju Alluri, and Anil Kumar Vuppala, "Toward improving the performance of epoch extraction from telephonic speech," *Circuits, Systems, and Sig. Process.*, pp. 1–15, 2020.

[18] P Gangamohan and S V Gangashetty, "Epoch extraction from speech signals using temporal and spectral cues by exploiting harmonic structure of impulse-like excitations," in *Proc. ICASSP*. IEEE, 2019, pp. 6505–6509.

[19] S R Kadiri, Alku Paavo, and B Yegnanarayana, "Comparison of glottal closure instants detection algorithms for emotional speech," in *Proc. ICASSP*. IEEE, 2020, pp. 7379–7383.

[20] A I Koutrouvelis, G P Kafentzis, N D Gaubitch, and R Heusdens, "A fast method for high-resolution voiced/unvoiced detection and glottal closure/opening instant estimation of speech," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 24, no. 2, pp. 316–328, 2015.

[21] V Khanagha, K Daoudi, and H M 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, 2014.

[22] W J Bogdan, "Saarbruecken voice database," 2007.

[23] American Speech-Language-Hearing Association, "Definitions of communication disorders and variations," 1993.

[24] Daniel R Boone, Stephen C McFarlane, Shelley L Von Berg, and Richard I Zraick, "The voice and voice therapy," 2005.

[25] J Baker, "Functional voice disorders: clinical presentations and differential diagnosis," in *Handbook of clinical neurology*, vol. 139, pp. 389–405. Elsevier, 2016.

[26] Eberhard Seifert and Juerg Kollbrunner, "An update in thinking about nonorganic voice disorders," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 132, no. 10, pp. 1128–1132, 2006.

[27] P Clarós, A Karlikowska, A Clarós-Pujol, A Clarós, and C Pujol, "Psychogenic voice disorders literature review, personal experiences with opera singers and case report of psychogenic dyspho-nia in opera singer," *Int J Depress Anxiety*, vol. 2, pp. 015, 2019.

[28] Krishna Gurugubelli and Anil Kumar Vuppala, "Stable implementation of zero frequency filtering of speech signals for efficient epoch extraction," *IEEE Sig. Process. Lett.*, vol. 26, no. 9, pp. 1310–1314, 2019.

[29] A Kounoudes, P A Naylor, and M Brookes, "The DYPSA algorithm for estimation of glottal closure instants in voiced speech," in *Proc. ICASSP*. IEEE, 2002, pp. 349–352.

[30] P A Naylor, A Kounoudes, J Gudnason, and M Brookes, "Estimation of glottal closure instants in voiced speech using the dypsa algorithm," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 15, no. 1, pp. 34–43, 2006.

[31] John Kane and Christer Gobl, "Evaluation of glottal closure instant detection in a range of voice qualities," *Speech Communication*, vol. 55, no. 2, pp. 295–314, 2013.