Stable Implementation of Zero Frequency Filtering of Speech Signals for Efficient Epoch Extraction

by

Krishna Gurugubelli, Anil Kumar Vuppala

in

IEEE Signal Processing Letters

Report No: IIIT/TR/2019/-1

Centre for Language Technologies Research Centre
International Institute of Information Technology
Hyderabad - 500 032, INDIA
September 2019

Stable Implementation of Zero Frequency Filtering of Speech Signals for Efficient Epoch Extraction

Krishna Gurugubelli and Anil Kumar Vuppala, *Member, IEEE*

***Abstract*—Epochs are the abrupt-closure events in vocal fold vibration during the production of voiced speech. Zero frequency filtering is a simple and effective technique used to estimate the glottal closure instants accurately from the speech signal. However, the zero frequency filter is an unstable system. Hence, it may not be suitable for practical implementation due to the requirements of high precision computation. In this letter, zero-phase zero frequency resonator is proposed as an alternative to zero frequency filter. The proposed approach provides a stable zero-phase response. The experimental results indicate that the performance of the proposed method outperformed the state-of-the-art methods in terms of identification rate 99.17% and provides comparable performance in terms of false alarm rate (0.41%), and identification accuracy (0.28 ms).**

***Index Terms*—Glottal closure instants, linear phase, stability, zero frequency filtering, zero phase.**

I. Introduction

THE instant of significant excitation due to the abrupt closure of the vocal folds during speech phonation is referred to as an epoch or glottal closure instant [1]. Determining the epoch locations from speech signal is useful in glottal source analysis [2], [3], glottal inverse filtering [4], text-to-speech synthesis [5], prosody modification [6], emotional speech analysis [7], and pathological speech analysis [8]. The state-of-the-art methods for epoch extraction are dynamic programming phase slope algorithm (DYPSA) [9], speech event detection using the residual excitation and a mean-based signal (SEDREAMS) method [10], yet another GCI algorithm (YAGA) [11], glottal closure/opening instant estimation forward-backward algorithm (GEFBA) [12], linear-prediction residual based methods [13], and zero frequency filtering [1].

The zero frequency filtering technique has been shown to be robust in the estimation of epoch locations. The fundamental idea behind the zero frequency filtering method is narrow-band filtering of the speech signal at zero frequency to obtain the evidence of epoch locations by minimizing the impact of time-varying vocal tract resonances [1]. The excellence of the zero frequency filtering lies in its simple design. However, the zero frequency filter (ZFF) is an unstable system as its response grows/decays with the polynomial degree of order three [14], [15]. Therefore, the practical implementation of ZFF may require higher precision. This work addresses the design aspects of zero frequency filtering such as stability, linear phase, and causality.

The present work is not the first in the literature, which attempts to improve the performance of the zero frequency filtering method. For example, in [14], [15], finite impulse response (FIR) implementations of zero frequency filtering are proposed to overcome the stability problem. In [15], an alternative method to zero frequency filtering is introduced to reduce the computational complexity by avoiding the trend removing steps. These methods are FIR filter approximations, which require a higher filter order. In [15], a filter order of 700 is used for the extraction of epoch locations. The proposed "zero-phase zero frequency resonator" (ZP-ZFR) is an infinite impulse response (IIR) filter approximation that requires lower filter order. The required filter order to meet a given specification is directly related to the hardware complexity, chip area, or computational speed of filter [16]. Hence, the proposed ZP-ZFR (4th order IIR filter) has a simple design, over the FIR approximation of ZFF. We hypothesized that ZP-ZFR guarantees the stability without phase distortion.

The rest of the letter is organized as follows: Section II provides a comprehensive analysis of zero frequency filtering. Section III derives the mathematical formulation of zero-phase zero frequency filtering. In Section IV, the efficiency of ZP-ZFR method is evaluated and compared with the baseline epoch extraction methods. The conclusive remarks are provided in Section V.

II. Analysis of Zero Frequency Filtering

This section discusses the frequency domain analysis of the ZFF. From [1], ZFF is a causal and IIR system having the transfer function $H_{ZFF}(z)$, and it is given by,

$$H_{ZFF}(z) = \frac{1}{(1-z^{-1})^4}. \tag{1}$$

The ZFF has four poles on the unit circle ($r = 1$). The frequency response of the ZFF is given by,

$$H_{ZFF}(e^{j\omega}) = \frac{1}{(1-e^{-j\omega})^4} \tag{2a}$$

$$= \frac{1}{(1-\cos(\omega)+j\sin(\omega))^4}. \tag{2b}$$

The magnitude response of ZFF is represented as,

$$|H_{ZFF}(\omega)| = \frac{1}{4(1-\cos(\omega))^2} \tag{3a}$$

Manuscript received April 15, 2019; revised June 24, 2019; accepted July 2, 2019. Date of publication July 23, 2019; date of current version July 30, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Tomoki Toda. *(Corresponding author: Krishna Gurugubelli.)*

The authors are with the Speech Processing Laboratory, LTRC, KCIS, International Institute of Information Technology, Hyderabad 500032, India (e-mail: krishna.gurugubelli@research.iiit.ac.in; anil.vuppala@iiit.ac.in).

Digital Object Identifier 10.1109/LSP.2019.2929442

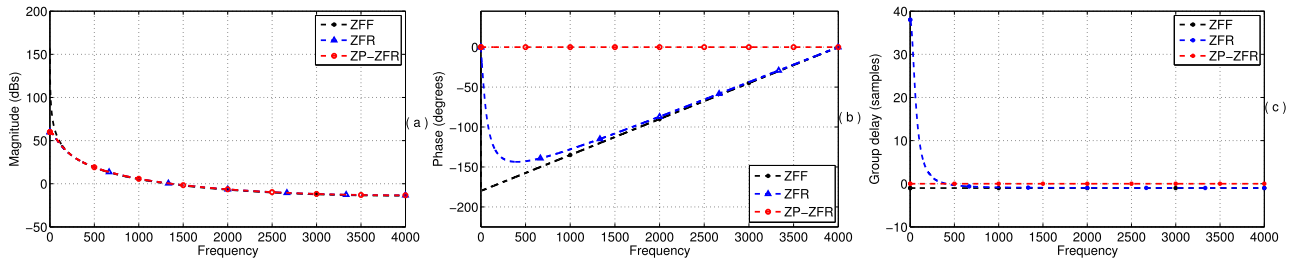

Fig. 1. The frequency response of zero frequency filter, zero frequency resonator (at $r = 0.98$) and zero-phase zero frequency resonator (at $r = 0.98$). (a) Magnitude response, (b) Phase response, and (c) Group-delay response.

$$= \frac{1}{16\sin^4(\omega/2)}. \tag{3b}$$

From equation (3b), it is observed that at $\omega = 0$, the magnitude response is unbounded. The ZFF has its maximum magnitude response at $\omega = 0$, and its magnitude response decays with increasing ω from 0 to $F_s/2$, as shown in Fig. 1(a). Here, F_s corresponds to the sampling rate. The phase response of the ZFF is given by,

$$\theta_{ZFF}(\omega) = -4\tan^{-1}\left(\frac{\sin(\omega)}{1-\cos(\omega)}\right) \tag{4a}$$

$$= 2(\omega - \pi). \tag{4b}$$

From equation (4b), it is observed that ZFF has linear phase response and constant group-delay of two samples. From the above discussion, it is concluded that ZFF is a causal, IIR filter having linear phase response and constant group-delay, i.e., no phase distortion. The same can be observed in Fig. 1(b) and Fig. 1(c). However, the repeated poles on the unit circle make the ZFF unstable. So, the response of the system may grow/decay rapidly. Due to the finite word length effects of digital filters, ZFF response may be stuck at saturation values. Consequently, the system may not be suitable in practice. The realization of a stable IIR filter having ZFF characteristics is discussed in the following section.

III. Formulation of Proposed Zero-Phase Zero Frequency Filtering

Consider the fourth order zero frequency resonator (ZFR) with poles located inside the unit circle represented by $H_{ZFR}(z)$. Its transfer function is given by,

$$H_{ZFR}(z) = \frac{1}{(1-rz^{-1})^4}. \tag{5}$$

Here r represents the pole location on z-plane, which is in the range of $0 < r < 1$. The frequency response of $H_{ZFR}(z)$ is given by,

$$H_{ZFR}(e^{j\omega}) = \frac{1}{(1-r\cos(\omega)+jr\sin(\omega))^4}. \tag{6}$$

The magnitude response of the ZFR is given by,

$$|H_{ZFR}(e^{j\omega})| = \frac{1}{((1-r\cos(\omega))^2+(r\sin(\omega))^2)^2} \tag{7a}$$

$$= \frac{1}{(1-2r\cos(\omega)+r^2)^2}. \tag{7b}$$

At $\omega = 0$, $H_{ZFR}(e^{j\omega})$ has finite magnitude response and it is equal to $\frac{1}{(1-r)^4}$. The value of r determines the bandwidth of the resonator. As r tends to 0, the magnitude response of the system at $\omega = 0$ becomes unity. For a given r value, the magnitude response decays as a function of frequency, as shown in Fig. 1(a). The phase response of the ZFR is given by,

$$\theta_{ZFR}(\omega) = -4\tan^{-1}\left(\frac{r\sin(\omega)}{1-r\cos(\omega)}\right). \tag{8}$$

The group-delay response of the ZFR is given by,

$$GD_{ZFR}(\omega) = \frac{2(r\cos(\omega)-r^2)}{1+r^2-2r\cos(\omega)}. \tag{9}$$

From equations (8) and (9), it is observed that the system has a non-linear phase response and non-linear group-delay response, as shown in Fig. 1(b) and Fig. 1(c) respectively. The non-linear group-delay of $H_{ZFR}(z)$ leads to phase distortion in the response of the system. The phase distortion may alter the location of the events like glottal closer instants in ZFR response. Due to the non-linear phase response of the ZFR system, it is not suitable for the epoch extraction as identification accuracy is important. The proposed method addresses stability, and linear phase issues are discussed in the following subsection.

A. Proposed Zero-Phase Zero Frequency Resonator

The proposed ZP-ZFR is obtained from the zero-phase implementation of a second-order zero frequency resonator $H(z)$. The transfer function of $H(z)$ is given by,

$$H(z) = \frac{1}{(1-rz^{-1})^2}. \tag{10}$$

The transfer function of the ZP-ZFR can be represented as,

$$H_{ZP-ZFR}(z) = H(z)H(z^{-1}) \tag{11a}$$

$$= \frac{1}{(1-rz^{-1})^2}\,\frac{1}{(1-rz)^2} \tag{11b}$$

$$= \frac{-z^{-2}}{(1-rz^{-1})^2(r-z^{-1})^2}. \tag{11c}$$

The frequency response of the ZP-ZFR is given by,

$$H_{ZP-ZFR}(e^{j\omega}) = H(e^{j\omega})H^*(e^{j\omega}) \tag{12a}$$

$$= \frac{1}{(1-re^{-j\omega})^2}\,\frac{1}{(1-re^{j\omega})^2} \tag{12b}$$

$$= \frac{1}{(1-2r\cos(\omega)+r^2)^2}. \tag{12c}$$

TABLE I
COMPARISON BETWEEN THE FILTER CHARACTERISTICS OF ZFF, ZFR, AND ZP-ZFR

Filter	Stability	Phase	Causality
ZFF	unstable	linear-phase	causal
ZFR	stable	nonlinear-phase	causal
ZP-ZFR	stable	zero-phase	non-causal

From equations (7b) and (12c), it can be observed that the magnitude response of ZP-ZFR is same as magnitude response of ZFR. The phase response of the proposed ZP-ZFR is equal to zero. So it has group-delay response $GD_{ZP-ZFR}(\omega)$ equal to zero samples ($GD_{ZP-ZFR}(\omega) = 0$). Hence, ZP-ZFR provides a stable response without any phase-distortion.

For $0 < r < 1$, ZP-ZFR has magnitude response equivalent to the magnitude response of ZFF, and it also has zero phase and group-delay responses as shown in Fig. 1(a-c). The proposed system has poles at $z = r$, and $1/r$. For the stability, the region of convergence (ROC) of the ZP-ZFR should include the unit circle. So the ROC of ZP-ZFR turns out to be an annular shape. By including the unit circle within it, and makes the whole system stable and non-causal.

Different properties of ZFF, ZFR, and ZP-ZFR are tabulated in Table I. The ZP-ZFR system has a stable, non-causal, zero-phase response. Though the ZP-ZFR is a non-causal system, it can be efficiently used for epoch extraction from the pre-recorded speech signals.

B. Choice of r Value and Trends in ZP-ZFR

The value of r in ZP-ZFR determines the bandwidth of the resonator. For lower values of r, the bandwidth of ZP-ZFR is very high. Therefore, the response of ZP-ZFR will have the higher order harmonics, which makes it difficult in finding the epoch locations due to the increased false alarms. For different values of r $(0.8,\ 0.85,\ 0.90,\ 0.95,\ 0.96,\ 0.97,\ 0.98,\ 0.99)$, the performance of ZP-ZFR is evaluated. For the values of r in between 0.95 and 0.99, the performance of the ZP-ZFR is found to be equivalent to ZFF. For the values of $r \leq 0.9$, the performance of ZP-ZFR is degraded due to the increase in false alarms.

The output of resonators for a typical speech signal is presented in Fig. 2. The polynomial decay (growth) can be observed in the response of ZFF, as shown in Fig. 2(a). The ZFR and ZP-ZFR exhibit oscillations in responses, as shown in Fig. 2(b) and Fig. 2(c), respectively. The oscillations in the responses of resonators correspond to very low-frequency components of speech, which are emphasized heavily due to the high gain of resonators around the zero frequency components of speech. It can be observed that the responses of ZFR and ZP-ZFR do not produce polynomial growth/decay, which is common in ZFF. Hence, high precision is not required in ZP-ZFR and ZFR methods. However, due to the very high gain around the zero frequency, the harmonics of speech are smeared over very low-frequency component, as shown in Fig. 2(b) and Fig. 2(c). Thus, ZP-ZFR based epoch extraction method also needs the trend removal operation. Generally, the fundamental period is in the range of 2.5 ms to 10 ms. By considering the highest fundamental period, the analysis window for trend removal should be greater than 10 ms, to avoid the false alarms. In this study, the length of the analysis window for trend removal is considered as 15 ms. In Fig. 3, the comparison between trend removal responses of resonators is demonstrated.

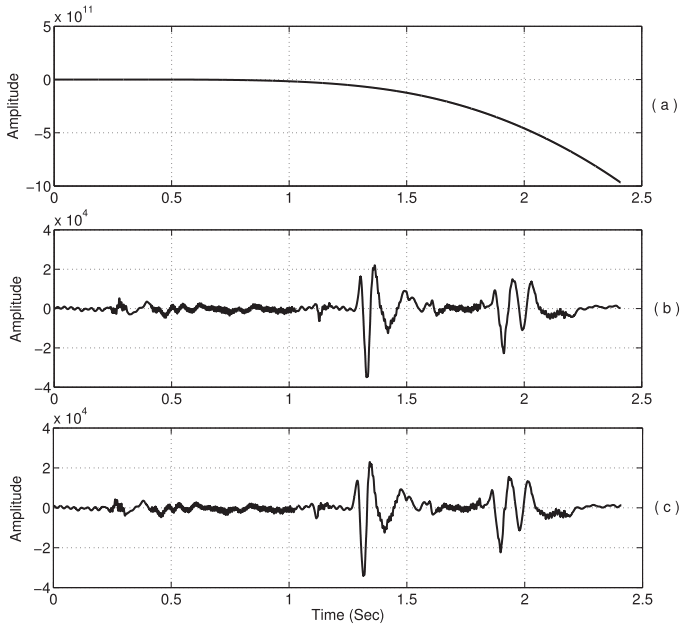

Fig. 2. Trends in the responses of resonators. (a) Response of the zero frequency filter, (b-c) Responses of zero frequency resonator, and zero-phase zero frequency resonator at r = 0.98 respectively.

From Fig. 3(b), it is understood that the ZFR response is delayed compared to ZFF response. The delay in ZFR response occurs due to its phase distortion. On the other hand, ZP-ZFR exhibits zero phase distortion. Consequently, the response of ZP-ZFR is aligned to the response of ZFF, as shown in Fig 3(c). Hence, it is concluded that the ZP-ZFR can be used for epoch extraction from a continuous speech signal. The set of steps involved in ZP-ZFR based epoch extraction mechanism is presented in the following subsection.

C. Steps for Extraction of Epochs Using ZP-ZFR

1) Pre-emphasize the speech signal.
2) The pre-emphasized speech signal is passed through the ZP-ZFR.
3) Apply the trend removal operation over the response of the ZP-ZFR, similar to the operation performed in ZFF [1]. In this work, the length of the trend removal window is considered as 15 ms.
4) Negative peaks in trend removed signal correspond to the epoch locations of a speech signal. Extraction of epoch locations from speech signal using the ZP-ZFR technique is demonstrated in Fig. 4.

IV. EVALUATION AND RESULTS

In this section, the proposed ZP-ZFR based epoch extraction method is compared with state-of-the-art epoch extraction methods on CMU-Arctic database [17]. The performance of the epoch extraction methods is evaluated using five standard measures namely identification rate (IDR in %), miss rate (MR in %), false alarm rate (FAR in %), and identification accuracy (IDA

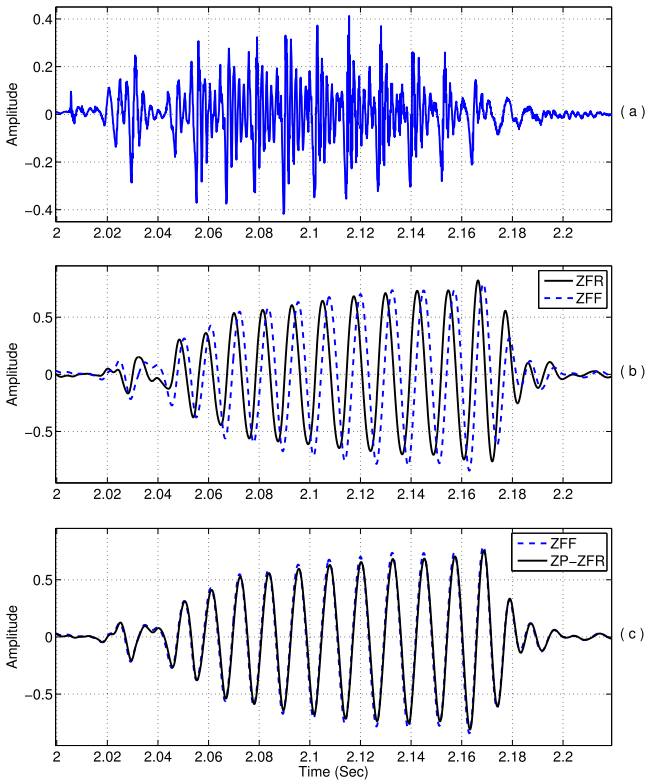

Fig. 3. Comparison between the trend removed responses of resonators. (a) Speech signal, (b–c) Trend removed responses of zero frequency resonator and zero-phase zero frequency resonator at r = 0.98. The dashed line in (b) and (c) represents the trend removed response of zero frequency filter.

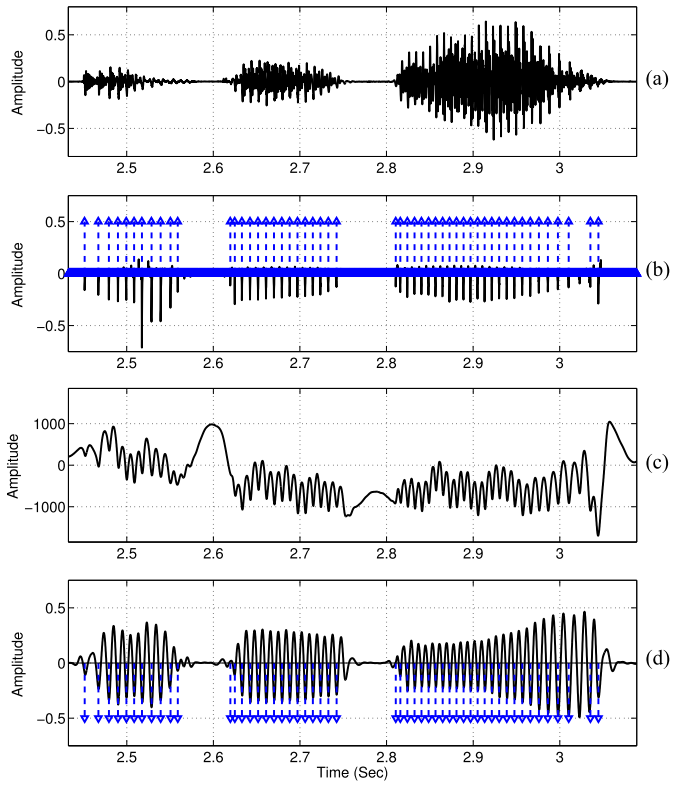

Fig. 4. ZP-ZFR based Epoch extraction. (a) Speech signal, (b) Derivative of the electroglottograph signal and corresponding ground truth epochs locations (blue spikes), (c) ZP-ZFR response, and (d) Trend removed ZP-ZFR response. Negative peaks of ZP-ZFR response are considered to be the epoch locations.

TABLE II
THE PERFORMANCE COMPARISON BETWEEN EPOCH EXTRACTION METHODS ON CMU-ARCTIC DATABASE. IDR–IDENTIFICATION RATE, MR–MISS RATE, FAR–FALSE ALARM RATE, AND IDA–IDENTIFICATION ACCURACY

Method	IDR (%)	MR (%)	FAR (%)	IDA (ms)
DYPSA	98.20	0.31	1.47	0.24
SEDREAMS	99.08	0.19	0.71	0.34
YAGA	99.03	0.08	0.88	**0.19**
SE-VQ	94.61	3.41	1.96	0.28
GEFBA	96.88	2.79	**0.33**	0.21
ZFF	98.72	**0.06**	1.20	0.26
ZP-ZFR	**99.17**	0.44	0.41	0.28

in ms). The detailed description of the evaluation measures can be found in [1].

Table II summarises the performance of the proposed ZP-ZFR method and the state-of-the-art methods DYPSA [9], SEDREAMS [10], YAGA [11], SE-VQ [18], GEFBA [12], and ZFF [1] for epoch extraction. From Table II, it is evident that ZFF, GEFBA, and YAGA methods perform better in terms of MR, FAR, and IDA, respectively. From the results, it can be seen that DYPSA and SEDREAMS methods have poor performance in terms of FAR and IDA, respectively. The SE-VQ method has the least performance in terms of IDR, MR, and FAR. The GEFBA method exhibits diminished performance in terms of IDR and MR. Though the performance of YAGA and DYPSA methods is comparable to baseline benchmark results, use of N-best dynamic programming makes these methods more complex than the simple signal processing techniques like GEFBA, ZFF, and ZP-ZFR. The ZFF method has relatively high FAR and provides an unstable response. Hence, this method may not be suitable in practice. The proposed ZP-ZFR efficiently approximates the zero frequency filtering and provides a stable response. Compared to the DYPSA, SEDREAMS, YAGA, and ZFF methods, the proposed ZP-ZFR method performs poor in terms of MR. However, the proposed ZP-ZFR method performs best in terms of IDR and exhibits comparable results in terms of IDA, and FAR.

V. CONCLUSION

In this letter, an efficient alternative approach for zero frequency filtering is proposed to find the epoch locations from speech signal. From the mathematical formulation, it is evident that the proposed zero-phase zero frequency resonator is a non-causal, stable, IIR system having a zero-phase response. Hence the proposed method can be efficiently realized on low-end computing devices. From the experimental results, it is evident that the proposed zero-phase zero frequency filtering method accurately estimates the epoch locations compared to the state-of-the-art methods in terms of identification rate (99.17%) and false alarm rate (0.41 %). In future work, we plan to investigate the robustness of the proposed method in different noisy conditions.

REFERENCES

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

[2] P. Alku, "Glottal inverse filtering analysis of human voice production—A review of estimation and parameterization methods of the glottal excitation and their applications," *Sadhana*, vol. 36, no. 5, pp. 623–650, 2011.

[3] B. R. Gerratt, J. Kreiman, and M. Garellek, "Comparing measures of voice quality from sustained phonation and continuous speech," *J. Speech, Lang., Hearing Res.*, vol. 59, no. 5, pp. 994–1001, 2016.

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

[5] J. P. Cabral, K. Richmond, J. Yamagishi, and S. Renals, "Glottal spectral separation for speech synthesis," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 2, pp. 195–208, Apr. 2014.

[6] K. S. Rao and B. Yegnanarayana, "Prosody modification using instants of significant excitation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 3, pp. 972–980, May 2006.

[7] S. M. Prasanna and D. Govind, "Analysis of excitation source information in emotional speech," in *Proc. INTERSPEECH*, 2010, pp. 781–784.

[8] P. Gómez-Vilda *et al.*, "Glottal source biometrical signature for voice pathology detection," *Speech Commun.*, vol. 51, no. 9, pp. 759–781, 2009.

[9] A. Kounoudes, P. A. Naylor, and M. Brookes, "The DYPSA algorithm for estimation of glottal closure instants in voiced speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, pp. 349–352.

[10] T. Drugman and T. Dutoit, "Glottal closure and opening instant detection from speech signals," in *Proc. INTERSPEECH*, 2009, pp. 2891–2894.

[11] M. R. Thomas, J. Gudnason, and P. A. Naylor, "Estimation of glottal closing and opening instants in voiced speech using the YAGA algorithm," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 82–91, Jan. 2012.

[12] 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. Audio, Speech, Lang. Process.*, vol. 24, no. 2, pp. 316–328, Feb. 2016.

[13] A. Prathosh, T. Ananthapadmanabha, and A. Ramakrishnan, "Epoch extraction based on integrated linear prediction residual using plosion index," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 12, pp. 2471–2480, Dec. 2013.

[14] K. S. Srinivas and K. Prahallad, "An FIR implementation of zero frequency filtering of speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 9, pp. 2613–2617, Nov. 2012.

[15] P. Gangamohan and B. Yegnanarayana, "A robust and alternative approach to zero frequency filtering method for epoch extraction," in *Proc. INTERSPEECH*, 2017, pp. 2297–2300.

[16] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 2009.

[17] J. Kominek and A. W. Black, "The CMU Arctic speech databases," in *Proc. 5th ISCA Speech Synthesis Workshop*, Pittsburgh, PA, USA, 2004, pp. 223–224.

[18] J. Kane and C. Gobl, "Evaluation of glottal closure instant detection in a range of voice qualities," *Speech Commun.*, vol. 55, no. 2, pp. 295–314, 2013.