Interspeech 2018
2-6 September 2018, Hyderabad

# Epoch Extraction from Pathological Children Speech Using Single Pole Filtering Approach

*C. M. Vikram*[1] *and S. R. Mahadeva Prasanna*[1,2]

[1]Indian Institute of Technology Guwahati, Guwahati, India
[2]Indian Institute of Technology Dharwad, Dharwad, India
{cmvikram, prasanna}@iitg.ernet.in

## Abstract

The instant of significant excitation of the vocal tract system is referred to the epoch of the speech signal. The presence of high pitch and aperiodicity are the major challenges for the epoch extraction from the speech of pathological children. In this work, impulse-like characteristics of epochs derived from single pole filter based time-frequency representation are exploited to propose an epoch extraction algorithm for the pathological children speech. The sharp transitions present in the single pole filtered envelope at the epochs are enhanced using multi-scale product computation. Further, the combined evidence derived from the multi-scale product of the filtered envelopes at different frequencies is used to locate the epochs. The proposed algorithm is evaluated over the Saarbruecken Voice Database containing pathological children speech and simultaneously recorded electroglottographic signals. The proposed method showed better identification accuracy for pathological children speech when compared to state-of-the-art techniques.

**Index Terms**: Pathological children speech, single pole filter, multi-scale product, epoch extraction.

## 1. Introduction

The primary excitation of vocal tract system takes place at the glottal closure instants (GCIs), which are referred to the epochs of voiced speech signal [1]. The knowledge of epochs is used for the computation of instantaneous fundamental frequency, the strength of glottal excitation, estimation of vocal tract spectral characteristics, and detection of voice disorders [2, 3, 4]. A detailed review of the importance of epoch-based analysis of speech is presented in Ref. [5]. Most of the existing epoch extraction algorithms work reasonably well for clean speech. By addressing the issues related to noisy speech, telephone quality speech, and emotional speech, alternative epoch extraction methods are proposed in Ref. [6, 7, 8]. But as per the knowledge of existing literature, the challenges related to epoch extraction from pathological children speech are not yet addressed. The presence of the high pitch and aperiodicity in the glottal vibrations [9] will pose challenges for the epoch extraction from pathological children speech.

Most of the existing epoch extraction algorithms, i.e., Hilbert Envelope (HE) based methods [10, 11], Dynamic Programming Phase Slope Algorithm (DYPSA) [12], Yet Another GCI / GOI Algorithm (YAGA) [13], and Dynamic Plosion Index (DPI) [14] depend on the residual or voice source signal derived from the linear prediction (LP) analysis of speech. The performance of these approaches critically relies on the accuracy of the vocal tract system modeled by LP analysis. LP analysis works better for low pitched speech, which is not suitable for the high pitched cases such as female and children speech [15]. In addition to high pitch, the presence of aperiodicity and glottal noise will increase the non-stationarity associated with the pathological children speech. The increased non-stationarity also creates a challenge for the LP-based epoch extraction methods.

Apart from LP-based methods, Zero Frequency Resonator (ZFR) and Speech Event Detection using the Residual Excitation And a Mean-based Signal (SEDREAMS) [6, 16] are proposed for the epoch extraction. These approaches involve the filtering of the speech signal at very low frequencies, where the influence of vocal tract resonances is considered to be minimum. ZFR showed robustness under noisy conditions in-terms of epoch detection rate than LP-based methods. However, ZFR is reported with poor identification accuracy than LP-based methods [17]. In SEDREAMS, the search intervals for the epochs are derived using mean based signal (low-pass filtered speech). Further, these search interval and LP residual are used to locate the epochs. Like ZFR, the smoothing operation in SEDREAMS improves the epoch identification rate under noisy conditions, and the usage of LP residual enhances epoch identification accuracy. However, LP analysis is not suitable for the pathological children speech to obtain accurate epoch locations. Therefore, epoch extraction from pathological children speech is still a challenge, and a reliable and accurate method is required for this.

Motivated by the importance of the epochs in speech analysis, in this work, an epoch extraction algorithm is proposed for the pathological children speech. The proposed epoch extraction method combines the advantages of epoch extraction algorithms developed for telephone quality speech [7] and emotional speech [8]. The time-frequency representation with better time-localized impulse-like events is derived using single pole filter (SPF) based approach [7]. Using SPF based time-frequency representation, the epochs are located by the combination of multi-scale product and zero-frequency filtering approach [8]. The proposed algorithm is evaluated and compared with state-of-the-art algorithms using Saarbruecken Voice Database [18]. The paper is organized as follows: Section 2 explains the proposed method for epoch extraction algorithm. Section 3 describes evaluation results on the database containing pathological children speech. Conclusion and possible future directions are mentioned in Section 4.

## 2. Proposed Epoch Extraction Algorithm

The proposed epoch extraction method consists of the estimation of filtered envelopes at different frequency bands using single pole filtering approach. SPF is a first order infinite impulse response (IIR) filter. The impulse response of first order IIR filter is exponentially decaying in nature, which can accurately characterize the sharp onsets present in the signal [19]. The advantages of SPF in the epoch extraction are explored

10.21437/Interspeech.2018-1613

in Ref. [7, 8]. The procedure for the extraction of SPF based time-frequency representation mentioned in Ref. [7] is briefly explained as follows.

Let $x(n)$ be the input signal, corresponding analytical signal $x_a(n)$ is computed using

$$x_a(n) = x(n) + jx_h(n) \tag{1}$$

where $x_h(n)$ is Hilbert transformed signal of $x(n)$. $x_a(n)$ is multiplied by a complex exponential of frequency $\pi - \omega_k$, i.e., $e^{\pi - \omega_k}$. This operation shifts the desired frequency components at $\omega_k$ to folding frequency $\pi$ $(f_s/2)$. This operation is equivalent in frequency domain as

$$X_{ak}(\omega) = X_a(\omega - (\pi - \omega_k)) \tag{2}$$

where $X_{ak}(\omega)$ and $X_a(\omega)$ are the spectra of $x_{ak}(n)$ and $x_a(n)$, respectively, and $\omega_k = \frac{2\pi f_k}{f_s}$. Here, $f_k$ is the frequency of desired spectral components need to be filtered and $f_s$ is the sampling frequency. The frequency translated signal $x_{ak}(n)$ is passed through SPF located at $\pi$. The transfer function ($H(z)$) and impulse response ($h(n)$) of SPF located at $\pi$ are given by

$$H(z)|_{\omega_k=\pi} = \frac{1}{1+rz^{-1}}, \quad h(n)|_{\omega_k=\pi} = (-r)^n u(n) \tag{3}$$

where $r$ is the radius of the pole. The pole is located at $z = -r$ in the z-plane with pole angle equals to $\pi$. The bandwidth $B$ of the frequency response and time constant $\tau$ of the time response of the filter are related to pole radius $r$ as

$$|r| = e^{-\pi BT}, \quad |r| = e^{-\frac{T}{\tau}} \tag{4}$$

where $T$ is the sampling interval. The output of the filter $y_k(n)$ is given by

$$y_k(n) = -ry_k(n-1) + x_k(n) \tag{5}$$

The envelope $e_k(n)$ of the filtered signal $y_k(n)$ is given by

$$e_k(n) = \sqrt{y_{kr}^2(n) + y_{ki}^2(n)} \tag{6}$$

where $y_{kr}(n)$ and $y_{ki}(n)$ are the real and imaginary components of $y_k(n)$, respectively. The desired frequency components around $f_k$ of $x_a(n)$ are shifted to $f_s/2$ and filtered by SPF. Thus, the envelope $e_k(n)$ corresponds to the envelope of the signal $x_k(n)$ filtered around a desired center frequency $f_k$. The value of $\tau$ is chosen based on the minimum pitch period of human, i.e., equal to 2 ms (corresponding to maximum pitch frequency of 500 Hz). Using SPF with $\tau = 2\,ms$, the envelopes ($e_k$) are computed by varying $f_k$ from 100 Hz to 7800 Hz, insteps of 10 Hz and time-frequency representation is obtained.

### 2.1. Comparison between SPF and SFF

The extraction of filtered envelopes using shifting and filtering at folding frequency ($f_s/2$) is proposed in Ref. [20], and the approach is named as single frequency filtering (SFF). The procedure for the estimation of time-frequency representation using SFF and SPF are appear to be similar. However, the methods differ in the choice of pole radius ($r$). In SFF, the $r$ value is almost equal to unity ($r = 0.99$), where the filter is highly narrow-band in nature. For the sampling frequency of 16000 Hz, the time constant $\tau$ corresponding to pole radius $r = 0.99$ is equal to 12.5 ms. Figures 1 (a)-(d), (e)-(h), and (i)-(l), represent the speech, differenced EGG (DEGG), SFF, and SPF based spectrograms for adult male, normal, and pathological children, respectively. The vertical striations of spectrogram corresponding to epochs are better visualized in SPF based spectrogram, when compared to SFF. Especially, for the pathological children speech case, the vertical striations are not prominent in SFF based spectrogram. This is due to the exponential smoothing effect caused by the filter. The degree of exponential smoothing is proportional to $\tau$. Larger $\tau$ associated with SFF will smooth the temporal transitions of speech present at the epochs. Therefore, the effect of strong excitation may mask the transitions due to weak excitations in SFF based spectrogram. But the smoothing effect is minimized in the case of SPF due to the reduction of filter's time-constant. Because, time-constant of SPF is chosen to equal to the minimum pitch period, which can resolve the sharp energy transitions at adjacent epochs. Therefore, impulse-like nature of epochs is better represented by the SPF based approach.

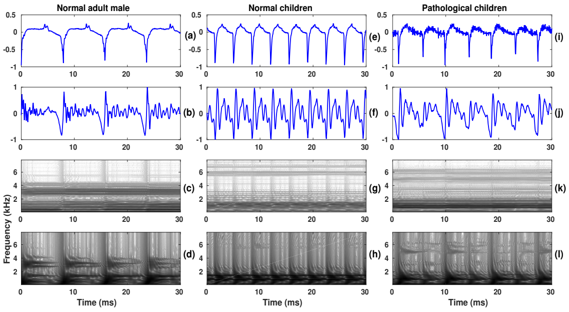

Figure 1: *(a)-(d), (e)-(h), and (i)-(l) represent DEGG, speech, SFF spectrogram (r=0.99), SPF spectrogram ($\tau$=2 ms, r=0.93 for $f_s$=16 kHz) for adult male, normal children, and pathological children, respectively.*

### 2.2. Multi-scale product of filtered envelopes

The epochs located by computing the time-marginal of SPF based time-frequency representation [7] showed poor identification accuracy (temporal delay in the epoch estimation). The combination of multi-scale product of SFF based filtered envelopes and zero-frequency filtered signal is used to locate the epochs [8]. However, the epoch information is better characterized in SPF based time-frequency representation, when compared to SFF. Therefore, in this work, multi-scale product based approach proposed in [8] is applied on SPF based time-frequency representation instead of SFF based approach. The procedure to compute the multi-scale product of SPF based filtered envelopes is explained as follows.

The multi-scale product of stationary wavelet transform (SWT) is used enhance the discontinuity information present in the signal [13]. SWT of signal $u'(n)$, $1 \leq n \leq M$ at scale $j$ is given by

$$d_j^s(n) = \sum_k g_j(k) a_{j-1}^s(n-k) \tag{7}$$

where $j = 1, ....J-1$ and the maximum scale $J$ is bounded by $log_2 M$. The coefficients $a_j^s$ are given by

$$a_j^s(n) = \sum_k h_j(k) a_{j-1}^s(n-k) \tag{8}$$

where $a_0^s = u^{'}(n)$. The $g_j(k)$ and $h_j(k)$ are detail and approximation filters, respectively. The multi-scale product is given by,

$$p(n) = \prod_{j=1}^{j_1} d_j(n) \tag{9}$$

where $j_1$ is chosen equal to 3 [13]. Let, the multi-scale product computed for $k^{th}$ filtered envelope $e_k(n)$ is denoted as $e_{pk}(n)$. Figures 2(a), (b), and (c) represent the DEGG, speech, and SPF based TFR, respectively. The multi-scale product of each filtered envelope is computed. The time-frequency representation derived using multi-scale product of filtered envelopes is shown in Figure 2(d). Vertical striations present in spectrogram (Figure 2(c)) corresponding to epochs are further highlighted, after computation of multi-scale product (Figure 2(d)). The $e_{pk}(n)$ computed at different frequencies are averaged to get the evidence $\beta(n)$. That is given by

$$\beta(n) = \frac{1}{N}\sum_{k=0}^{M-1} e_{pk}(n) \tag{10}$$

where $M$ is the number of filtered envelopes. The plot of $\beta(n)$ is shown in Figure 2(f), which shows close resemblance with DEGG (Figure 2(a)). Table. 1 shows the normalized cross-correlation coefficients computed for DEGG vs. ILPR, and DEGG vs. $\beta(n)$ for BDL male, SLT female, and children speakers. Samples of BDL and SLT speakers are obtained from CMU arctic database [21], whereas children samples are taken from Saarbruecken database [18]. The normalized cross-correlation coefficient values in Table 1 indicates that the correlation values are high for $\beta(n)$, when compared to ILPR. Generally, the inverse filtered signal is believed to have high correlation with DEGG signal. But the correlation values indicates that a non-LP based approach, i.e., SPF based approach can also better characterizes the glottal source information, than LP inverse-filter based approach.

Table 1: *Normalized cross correlation coefficients between source representative signals and DEGG*

| Method | Datasets and cross correlation values | | |
|---|---|---|---|
| | Adult male<br>(BDL) | Adult female<br>(SLT) | Pathological<br>children |
| ILPR | 0.40 | 0.39 | 0.57 |
| MSP-SPF | 0.48 | 0.59 | 0.61 |

### 2.3. Procedure to locate the epochs

Within each glottal cycle, the strongest negative peak of $\beta(n)$ coincides with the negative peak of DEGG. Similar to epoch extraction procedure mentioned in Ref. [8], in this work, the combination of zero frequency filtered signal (ZFFS) and SPF based approach is used to locate the epochs. First, the positive zero crossings of ZFFS are obtained. Around the positive zero crossings, a search interval of 1.2 ms is considered. Within this search interval, the location of the strongest negative peak of $\beta(n)$ is considered as the epoch location. Figure 2(e) shows that ZFFS along with the reference interval marked around each positive zero crossings. The detected epochs using $\beta(n)$ are shown in Figure 2(f). The proposed epoch extraction approach is based on SPF and multi-scale product (MSP), which is referred SPF-MSP method.

## 3. Evaluation and Results

The proposed epoch extraction algorithm is evaluated on Saarbruecken pathological voice database [18]. The database consists of simultaneously recorded speech and EGG signals of adult and children speakers. The database includes the speech and EGG samples of different pathologies such as the cyst, polyp, dysphonia,...,etc. The recordings consist of vowels /a/, /i/, and /u/ in four different conditions: neutral, increasing, decreasing pitch, and the combination of increasing and decreasing pitch. Also the combination of three vowels (/a/, /i/, and /u/) and sentences are present in the database. In this work, 184 utterances of 13 children with voice pathologies belonging to age group of 5-15 years are considered from Saarbruecken database. The speech and EGG samples present in the database are recorded at 50000 Hz sampling rate. Further, the samples are down-sampled to 16000 Hz.

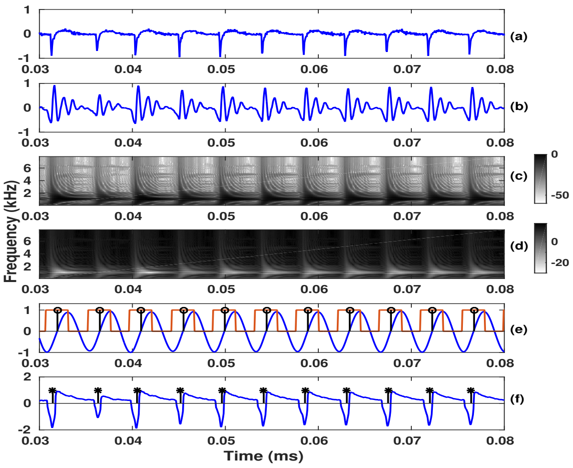

Figure 2: *(a) DEGG, (b) speech, (c) SPF spectrogram, (d) 2-D representation of multi-scale product of filtered envelopes, (e) ZFFS (blue color) with search interval (red color) around each positive zero crossings, and (f) average of multi-scale product of filtered envelopes ($\beta(n)$) with detected epochs.*

Table 2: *Performance of Epoch Extraction on Pathological Children Speech Database (Saarbruecken database [18])*

| Method | IDR (%) | MR (%) | FAR (%) | IDA (ms) | IDR<br>± 25 ms (%) |
|---|---|---|---|---|---|
| DYPSA | 93.37 | 5.4 | 0.82 | 0.61 | 51.16 |
| YAGA | 91.69 | 3.64 | 4.25 | 0.9 | 35.46 |
| DPI | 92.64 | 6.23 | 0.72 | 0.57 | 54.06 |
| ZFR | **96.7** | 2.25 | **0.63** | 0.54 | 54.81 |
| SEDREAMS | 95.12 | 2.6 | 1.87 | 0.62 | 53.83 |
| SPF+TM | 92.49 | 5.01 | 2.09 | 0.74 | 37.66 |
| SPF+MSP ($\tau = 2\, ms$) | 96.47 | **2.18** | 0.94 | **0.27** | **83.17** |
| SFF+MSP ($r = 0.99$) | 96.42 | 2.18 | 0.98 | 0.42 | 76.97 |

### 3.1. Ground truth

In the literature, the ground truth epoch locations are derived from DEGG either by peak picking method or sigma algorithm Ref. [22]. For pathological children speech, the amplitude of peaks corresponding to epochs will vary dynamically. Hence, ground truth epoch locations are derived using threshold-based peak picking from DEGG may not be appropriate for pathological speech case. Automatic detection of ground truth epoch locations from EGG signal using sigma algorithm is proposed in Ref. [22]. Sigma algorithm uses dynamic programming algorithm to locate the ground truth epochs from the EGG, where the parameters are set for normal adult speakers. These parameters may not be suitable for the EGG of pathological children. In this work, a semi-automatic method is implemented to derive the ground truth epoch locations from DEGG. First DC drift of EGG is removed using a butter worth high pass filter with the cut-off frequency equal 10 Hz. The high-pass filtered EGG is differenced to get DEGG. The ZFFS of DEGG is obtained, and the positive zero crossings are determined. The ground truth epoch locations are obtained by picking largest negative peak of DEGG around the positive zero crossing of ZFFS. Further, the detected epoch locations are visually verified using Wavesurfer tool [23]. Falsely identified epochs are removed, and mis de-

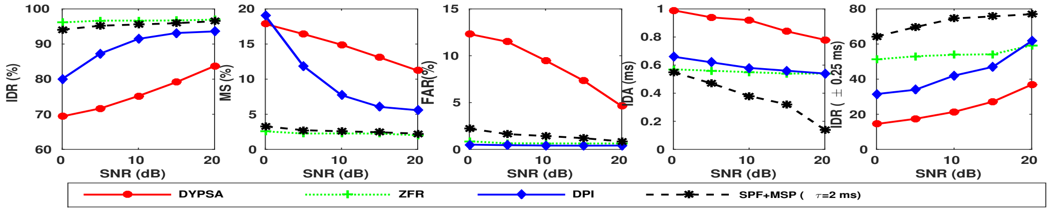

Figure 3: *Robustness against white noise. Figure shows the performance of four different epoch extraction algorithms on pathological children speech data at different SNRs (0 to 20 dB).*

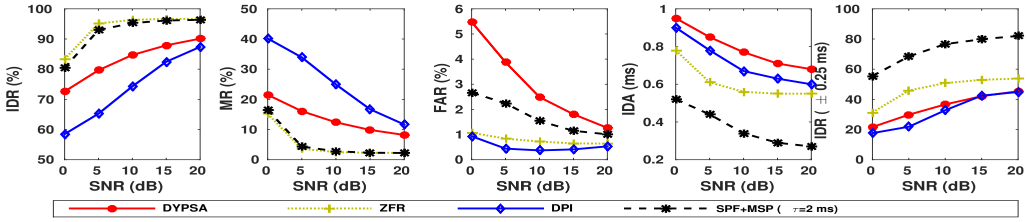

Figure 4: *Robustness against babble noise. Figure shows the performance of four different epoch extraction algorithms on pathological children speech data at different SNRs (0 to 20 dB).*

tected epochs are again marked using Wavesurfer tool. The time delay between EGG and speech recording varies across speakers due to the variability in the mouth-to-microphone distance. The time-delay between EGG and speech is adjusted for each utterance using the procedure mentioned in Ref. [15].

### 3.2. Results and Discussion

The proposed SPF-MSP based epoch extraction algorithm is evaluated using parameters: identification rate (IDR), mis rate (MR), false alarm rate (FAR), identification accuracy (IDA), and IDR within ± 0.25 ms. The IDR, MR, and FAR represent the epoch detection performance, whereas IDA and IDR within ± 0.25 ms give the temporal accuracy of the algorithm [17]. LP-based approaches: DYPSA, DPI, and YAGA, smoothing approaches: ZFR and SEDREAMS, SPF based algorithm using time marginal (SPF-TM) [7] and SFF based multi-scale product approach (SFF-MSP) [8] are considered as the state-of-the-art methods. The evaluation results are presented in Table 2. The LP-based methods (DYPSA, DPI, and YAGA) showed poor performance when compared to non-LP based approaches (ZFR, SEDREAMS, SPF-MSP, and SFF-MSP). This indicates that the presence of high pitch and aperiodicity in the pathological children speech affect the performance of LP-based approaches. The ZFR showed better IDR, but the performance in terms of accuracy parameters, i.e., IDA and IDR within±0.25 ms are noticed to be poor. The IDR of proposed SPF-MSP is comparable with that of ZFR, whereas IDA and IDR within±0.25 ms are better than that of ZFR. The peaks of LP residual are used to locate the epochs by SEDREAMS algorithm, where the reference interval derived using mean based signal is used to search the peaks. As the LP analysis is not reliable for high pitched children speech, SEDREAMS results in poor IDA.

The SPF-TM algorithm resulted in poor IDA due to the delay introduced by the vocal tract resonances[7]. In SFF-MSP, the filter is implemented with very larger time-constant. Therefore, SFF-MSP gives greater IDA and lesser IDR within ±0.25 ms when compared to the proposed SPF-MSP method. This indicates that increase in the temporal resolution by the reduction of $r$ value in SPF-MSP approach improves the temporal accuracy of detected epochs.

### 3.3. Robustness to Noisy Conditions

The robustness of DYPSA, DPI, ZFR, and proposed SPF-MSP algorithms against the white and babble noise conditions is evaluated. The noise samples are taken from NOISEX database [24] and added to pathological children speech samples of Saarbruecken pathological voice database. Figures. 3 and 4 show the IDR, MR, FAR, IDA and IDR within ± 0.25 ms of different for the white and babble noise cases, respectively. The proposed algorithm primarily locates the epochs using ZFR. Further, ZFR based epoch locations are refined using SPF-MSP based impulse-like sequence. Hence, the proposed method shows better robustness to noisy cases, which is almost equal to ZFR. Also, under the noisy conditions, the SPF-MSP shows better IDA and IDR within ± 0.25 ms than other methods. Thus the combination of ZFR and SPF based approaches showed robustness to noisy cases, in-terms of both identification rate and identification accuracy.

## 4. Conclusion and Future Work

In this work, an epoch extraction algorithm for the pathological children speech is proposed. The impulse-like sequence derived using the SPF based envelopes and the multi-scale product is used as the pre-processed signal for the epoch extraction. A short reference interval defined around the positive zero crossings of ZFFS is used to locate the epoch from the pre-processed signal. The proposed approach showed better performance for the pathological children speech database, under clean as well noisy conditions, when compared to LP-based approaches. The future work includes the application of proposed epoch extraction algorithm for the epoch-synchronous analysis of various voice disorders.

## 5. Acknowledgement

This work is in part supported by the project grants, for the projects entitled “NASOSPEECH: Development of Diagnostic system for Severity Assessment of the Disordered Speech” funded by the Department of Biotechnology (DBT), Govt. of India and “ARTICULATE +: A system for automated assessment and rehabilitation of persons with articulation disorders” funded by the Ministry of Human Resource Development (MHRD), Govt. of India.

# 6. References

[1] T. V. Ananthapadmanabha and B. Yegnanarayana, “Epoch extraction of voiced speech,” *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 23, no. 6, pp. 562–570, 1975.

[2] B. Yegnanarayana and R. N. Veldhuis, “Extraction of vocal-tract system characteristics from speech signals,” *IEEE trans. on Speech and Audio Process.*, vol. 6, no. 4, pp. 313–327, 1998.

[3] B. Yegnanarayana and K. S. R. Murty, “Event-based instantaneous fundamental frequency estimation from speech signals,” *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 17, no. 4, pp. 614–624, 2009.

[4] N. Adiga, C. M. Vikram, K. Pullela, and S. R. M. Prasanna, “Zero frequency filter based analysis of voice disorders,” 2017.

[5] B. Yegnanarayana and S. V. Gangashetty, “Epoch-based analysis of speech signals,” *Sadhana*, vol. 36, no. 5, pp. 651–697, 2011.

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

[7] C. M. Vikram and S. R. M. Prasanna, “Epoch extraction from telephone quality speech using single pole filter,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 3, pp. 624–636, 2017.

[8] S. R. Kadiri and B. Yegnanarayana, “Epoch extraction from emotional speech using single frequency filtering approach,” *Speech Communication*, vol. 86, pp. 52–63, 2017.

[9] J. Kreiman and B. R. Gerratt, “Perception of aperiodicity in pathological voice,” *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2201–2211, 2005.

[10] T. V. Ananthapadmanabha and B. Yegnanarayana, “Epoch extraction from linear prediction residual for identification of closed glottis interval,” *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 27, no. 4, pp. 309–319, 1979.

[11] 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, 2007.

[12] A. Kounoudes, P. A. Naylor, and M. Brookes, “The dypsa algorithm for estimation of glottal closure instants in voiced speech,” in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 1. IEEE, 2002, pp. I–349.

[13] 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, and Lang. Process.*, vol. 20, no. 1, pp. 82–91, 2012.

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

[15] O. Babacan, T. Drugman, N. d’Alessandro, N. Henrich, and T. Dutoit, “A quantitative comparison of glottal closure instant estimation algorithmson a large variety of singing sounds,” in *14th Annual Conference of the International Speech Communication Association (Interspeech 2013)*, 2013, pp. 1–5.

[16] T. Drugman and T. Dutoit, “Glottal closure and opening instant detection from speech signals.” in *Proc. of Interspeech*, 2009, pp. 2891–2894.

[17] T. Drugman, M. Thomas, J. Gudnason, P. Naylor, and T. Dutoit, “Detection of glottal closure instants from speech signals: a quantitative review,” *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 20, no. 3, pp. 994–1006, 2012.

[18] B. Woldert-Jokisz, “Saarbruecken voice database,” 2007.

[19] S. Tomazic, “On short-time fourier transform with single-sided exponential window,” *Signal processing*, vol. 55, no. 2, pp. 141–148, 1996.

[20] G. Aneeja and B. Yegnanarayana, “Single frequency filtering approach for discriminating speech and nonspeech,” *IEEE/ACM Trans. Audio, Speech, and Lang. Process.*, vol. 23, no. 4, pp. 705–717, 2015.

[21] “The Festvox Website.” [Online]. Available: http://festvox.org

[22] M. R. Thomas and P. A. Naylor, “The sigma algorithm: A glottal activity detector for electroglottographic signals,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1557–1566, 2009.

[23] K. Sjölander and J. Beskow, “Wavesurfer-an open source speech tool,” in *Sixth International Conference on Spoken Language Processing*, 2000.

[24] “Noisex-92.” [Online]. Available: http://www.speech.cs.cmu.edu/comp.speech/Sectionl/Data/-noisex.html