Interspeech 2018
2-6 September 2018, Hyderabad

Detection of Glottal Excitation Epochs in Speech Signal Using Hilbert Envelope

Hirak Dasgupta, Prem C. Pandey, K. S. Nataraj

Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
hirakdgpt@ee.iitb.ac.in, pcpandey@ee.iitb.ac.in, natarajks@ee.iitb.ac.in

Abstract

A technique, suitable for real-time processing, is presented for detection of glottal excitation epochs in voiced speech. It uses Hilbert envelope to enhance saliency of the glottal excitation epochs and to reduce the ripples due to the vocal tract filter. The processing comprises the steps of dynamic range compression, calculation of the Hilbert envelope, and epoch marking. The first step reduces amplitude variation by applying A-law on the signal envelope. The second step calculates the Hilbert envelope using the output of an FIR filter-based Hilbert transformer and the delay-compensated signal. The third step uses a dynamic peak detector with fast rise and slow fall and nonlinear smoothing using a two-step median-mean filter to further enhance the saliency of the epochs, followed by a differentiator to mark them. The technique is tested using the CMU-ARCTIC database with simultaneously recorded speech and EGG signals. The results showed a good match in the performance of the proposed technique with those of the state-of-the-art techniques and its robustness against highpass filtering. It may be useful for diagnosis of voice disorders and high-quality voice conversion.

Index Terms: fundamental frequency, glottal excitation epoch, Hilbert envelope, pitch period

1. Introduction

Voiced speech is the output of time-varying vocal tract filter excited by pulsatile airflow due to quasi-periodic vibration of the glottal folds [1]. The excitation is characterized by an impulsive excitation around the instants of glottal closure, known as the excitation epochs [2] and the duration between two successive epochs is termed as the pitch period. Pitch estimation methods can be categorized as window-based or event-based. The window-based methods treat the speech signal as stationary for the window duration and hence cannot track fast changes in the pitch. The event-based methods locate points associated with a significant event or phase in each cycle of the excitation. Epoch detection is useful in many speech processing applications such as de-reverberation of speech signal [3], diagnosing disorders of the vocal folds [4]-[7], high-quality voice conversion [8], and for accurate estimation of the vocal tract filter response [9] etc.

Several event-based techniques [10]-[23] have been reported for epoch detection of the speech signal. In [10], the vocal tract response is reduced by passing the pre-emphasized speech signal through two marginally stable cascaded zero frequency resonators (ZFR). The positive zero-crossings of the sinusoid-like signal generated by repeated mean-subtraction operation of the output of the resonator represent the glottal closure instants (GCIs). In the technique named as 'speech event detection using the residual excitation and the mean based signal' (SEDREAMS) [11], the epoch containing intervals are marked from the local-minima to the subsequent positive zero-crossings on a running mean-based speech signal and the highest peaks of the LP residual in these intervals are marked as the epochs. Patil and Viswanath [12] and Shikhah and Deriche [13] used Teager energy operator on a lowpass filtered speech for GCI detection. In [15]-[16], Hilbert envelope of the linear prediction (LP) residual is used to detect the epochs. In [17], integrated LP residual (ILPR) is calculated and the modified short-time crest factor of the half-wave rectified ILPR, termed as the dynamic plosion index, is used to detect the GCIs. In [18], a recursive algorithm on a temporal measure termed as the cumulative impulse strength derived from the ILPR is used for GCI detection. In [19], the epochs are marked as the positive zero-crossings of the average phase slope function of the unwrapped phase spectrum of the LP residual. In the dynamic programming phase slope algorithm (DYPSA) [20], the instants of significant excitation are detected by selecting the best possible set of epochs from the initial hypothesized points calculated using an energy-weighted group delay function and a phase slope projection method. Vikram and Prasanna [21] detected epochs of telephony speech by localizing vertical striations in time-frequency representation of voiced speech using a single-pole filter based filter bank approach.

Epoch detection techniques based on the computation of the LP residual suffer from error due to the inaccurate modeling of the vocal tract transfer function and bipolar swing around the epochs due to the phase angle of formants [23]. The ZFR and SEDREAMS techniques require the presence of the fundamental and hence cannot be used for epoch detection of highpass filtered speech. For most real-time applications, the processing should involve single-pass operations with a total delay (sum of algorithmic and computational delays) of less than 125 ms, the detectability threshold for audio-visual delay [24]. Here we present a new technique for epoch detection, which is suitable for real-time processing and is robust against highpass filtering. It uses the Hilbert envelope of the speech signal to enhance the excitation epochs and to suppress the ripples related to the vocal tract response, a dynamic peak detector with fast rise and slow fall along with a nonlinear smoother to further enhance the saliency of the epochs, and a differentiation-based saliency to mark them.

The basis for the proposed technique and its implementation are described in the second and third sections, respectively. The test results are presented in the fourth section, followed by conclusion in the last section.

2. Basis for the proposed technique

Speech signal during the voiced segments is modeled as the convolution of the impulse response of the time-varying vocal tract and glottal filters and a quasi-periodic impulse train as the excitation. The voiced speech signal $s(n)$ can be represented using its short-time harmonic model as

10.21437/Interspeech.2018-2014

$$s(n) = \sum_{k=1,N} b_k \cos(k\omega_0 n + \theta_k) \quad (1)$$

where b_k and θ_k represent the combined effect of the vocal tract and glottal filters and ω_0 is the fundamental frequency.

The Hilbert envelope of $s(n)$ is the magnitude of the complex analytic signal $s_a(n) = s(n) + js_h(n)$, where $s_h(n)$ is the Hilbert transform of $s(n)$ and can be obtained by a $\pi/2$-phase shifter, also known as the Hilbert transformer [25], with the frequency and impulse responses given as

$$H(\omega) = \begin{cases} -j, & 0 < \omega < \pi \\ 0, & \omega = 0, \ \pi \\ j, & -\pi < \omega < 0 \end{cases} \quad (2)$$

$$h(n) = \begin{cases} \sin^2(n\pi/2)/(n\pi/2), & n \neq 0 \\ 0 & n = 0 \end{cases} \quad (3)$$

The square of the Hilbert envelope is given as

$$e_h(n) = s^2(n) + s_h^2(n) \quad (4)$$

For speech signal $s(n)$ in (1), $s_h(n)$ can be given as

$$s_h(n) = \sum_{k=1,N} b_k \sin(k\omega_0 n + \theta_k) \quad (5)$$

and the square of the Hilbert envelope can be given as

$$\begin{aligned} e_h(n) = & \sum_{q=1}^{N} b_q^2 + 2\sum_{q=1}^{N-1} b_q b_{q+1} \cos(\omega_0 n + \theta_{q+1} - \theta_q) \\ & + 2\sum_{q=1}^{N-2} b_q b_{q+2} \cos(2\omega_0 n + \theta_{q+2} - \theta_q) + \cdots \\ & + 2b_1 b_N \cos\{(N-1)\omega_0 n + \theta_N - \theta_1\} \end{aligned} \quad (6)$$

The envelope consists of an offset and sum of harmonics of ω_0, with several harmonics in $s(n)$ contributing to the fundamental and enhancing the instants of significant excitation. Examples of the Hilbert envelope for speech waveforms in Figure 1 show enhancement of periodic excitation in case of vowels and also in case of highpass filtered vowels.

3. Proposed epoch detector

The proposed technique is shown in Figure 2, with the processing blocks of dynamic range compression, Hilbert envelope calculation, and epoch marking. Dynamic range compression acts as a pre-processing step before the Hilbert envelope calculation to reduce the possibility of misdetection of the epochs during low-level segments. Hilbert envelope is calculated using a Hilbert transformer realized as an FIR filter. The epoch marking uses a dynamic peak detector followed by nonlinear smoother to further reduce the residual ripples in the Hilbert envelope without reducing the saliency of the epochs and it uses a differentiator-based saliency detector to mark the epochs. The three blocks are devised for making the technique suitable for applications requiring real-time processing, with single-pass operations and total algorithmic delay much below 125 ms. The blocks are further described in the following subsections, with the values of the processing parameters given for sampling frequency of 10 kHz.

3.1. Dynamic range compression

Dynamic range compression (DRC) is implemented by applying feed-forward compression, based on the A-law [26], on the envelope of the input signal $s_{in}(n)$, as shown in Figure 3(a). The envelope $a(n)$ is calculated as the short-time average magnitude of the signal using the following recursive equation:

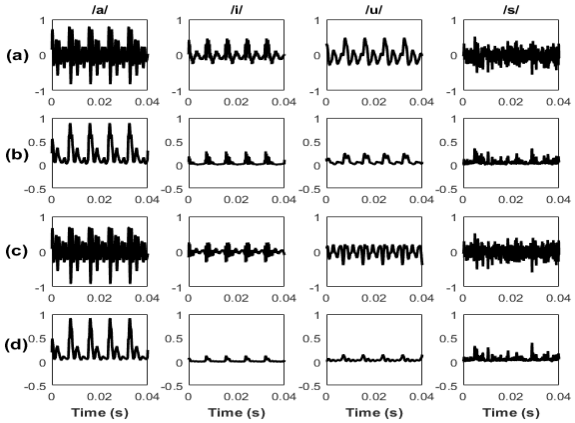

Figure 1: *Hilbert envelope examples: (a) waveforms of three synthesized vowels (120 Hz pitch) and a fricative, (b) Hilbert envelope of the waveforms in (a), (c) high–pass filtered (300 Hz cutoff) waveforms corresponding to (a), (d) Hilbert envelope of the waveforms in (c).*

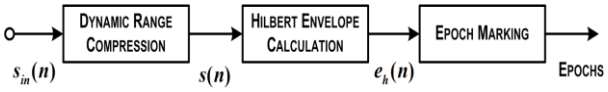

Figure 2: *Proposed epoch detector.*

$$a(n) = a(n-1) + [|s_{in}(n)| - |s_{in}(n-L)|]/L \quad (7)$$

where L corresponds to a 25-ms window. For input signal range of [-1,+1], the A-law compressed envelope is given as

$$\tilde{a}(n) = \begin{cases} Aa(n)/(1+\ln A), & 0 \le a(n) \le 1/A \\ [1+\ln\{Aa(n)\}]/(1+\ln A), & 1/A < a(n) \le 1 \end{cases} \quad (8)$$

The compressed signal $s(n)$ is obtained by multiplying the input signal, with the delay equal to that in the envelope calculation, with a time-varying scaling factor as

$$s(n) = [\tilde{a}(n)/a(n)]\, s_{in}(n-(L-1)/2) \quad (9)$$

The value of A in (8) is set as 40 to provide compression without excessive increase of noise during the silences and it results in the highest gain of approximately 19 dB. Figure 3(b) shows the variation of A-law compressed envelope $\tilde{a}$ with the signal envelope a.

3.2. Hilbert envelope calculation

The Hilbert transform of the signal is obtained using an FIR filter with impulse response obtained by applying a Hamming window of length M on the non-causal impulse response of the Hilbert transformer as given in (3) and $(M-1)/2$-sample shift. The envelope $e_{ht}(n)$ is calculated, as shown in Figure 4, from the output of the Hilbert transformer $s_{ht}(n)$ and the delay-compensated input $s_d(n)$ using the following equations:

$$s_{ht}(n) = s(n) * h(n) \quad (10)$$

$$s_d(n) = s(n-(M-1)/2) \quad (11)$$

$$e_{ht}(n) = s_{ht}^2(n) + s_d^2(n) \quad (12)$$

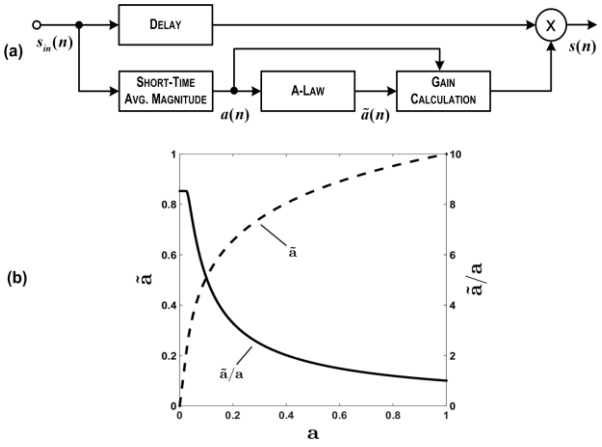

Figure 3: *DRC using A-law based feed-forward compression of the envelope: (a) block diagram of implementation, (b) variation of* $\tilde{a}$ *and* $\tilde{a}/a$ *with* a **.**

To enable suppression of glottal and vocal tract filter responses without excessive smearing of the representation of the glottal excitation in the envelope, *M* is empirically selected to correspond to 15 ms. The algorithmic delay in obtaining $e_h(n)$ is $(L+M-2)/2$ samples. With the values of *L* and *M* as selected here, this delay is 20 ms.

3.3. Epoch marking

The epoch marking block comprises a dynamic peak detector followed by nonlinear smoother and a differentiation-based saliency detector.

The peak detector is realized for updating peak $c(n)$ and valley $d(n)$ using the following recursive equations,

$$c(n)=\begin{cases}\mu c(n-1)+(1-\mu)e_{ht}(n), & \text{if } e_{ht}(n)\geq c(n-1)\\ \nu c(n-1)+(1-\nu)d(n-1), & \text{otherwise}\end{cases} \quad (13)$$

$$d(n)=\begin{cases}\mu d(n-1)+(1-\mu)e_{ht}(n), & \text{if } e_{ht}(n)\leq d(n-1)\\ \nu d(n-1)+(1-\nu)c(n-1), & \text{otherwise}\end{cases} \quad (14)$$

The peak $c(n)$ falls asymptotically to $d(n)$, which tracks the offset in the Hilbert envelope. The rise and fall rates are controlled by the constants μ and ν, selected to be in the range [0,1]. A fast rise (small μ) and slow fall (large ν) help in suppressing the ripples while retaining saliency of the epochs. We have used μ = 0.1 and ν = 0.9954 for 90% rise in one sample and 60% fall in 100 samples.

A nonlinear smoothing, comprising a two-step median-mean filter [27], as shown in Figure 5, is used to further suppress the residual ripples in the peak detector output. The first median-mean filter reduces the small ripples without smearing the large transitions and the second median-mean filter helps in restoring the peak-valley contrast. An 11-point median and 3-point mean filter was found to be effective for ripple suppression without disturbing epoch saliency.

The output $x(n)$ of the nonlinear smoother is used for locating the salient points related to the instants of glottal excitation. Differentiation is carried out using the following 5-point difference equation:

$$y(n)=[-x(n)+8x(n-1)-8x(n-3)+x(n-4)]/12 \quad (15)$$

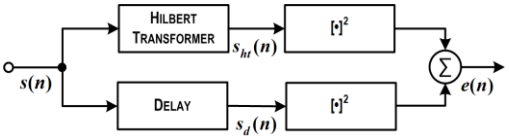

Figure 4: *Implementation of Hilbert envelope using FIR filter-based Hilbert transformer.*

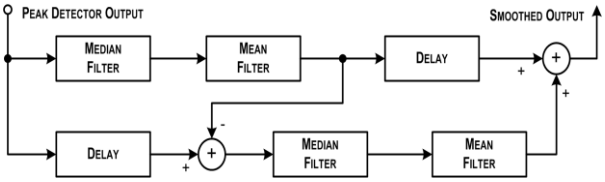

Figure 5: *Nonlinear smoother using two-step median-mean filter.*

The salient points corresponding to the excitation impulses are detected by applying an amplitude-duration thresholding on $y(n)$. The amplitude threshold $A_\theta(n)$ is calculated as the short-time average magnitude of the differentiator output as

$$A_\theta(n)=A_\theta(n-1)+[|y(n)|-|y(n-P)|]/P \quad (16)$$

where *P* corresponds to a 10-ms window. The duration threshold $T_\theta(n)$ is calculated as half of the mean of the previous 10 pitch periods and lying within 2 – 10 ms. A point is marked as an epoch if $y(n)$ exceeds $A_\theta(n)$ and the time difference between this point and the last detected epoch exceeds $T_\theta(n)$ .The initial value for $T_\theta(n)$ is set as 2 ms.

4. Implementation and evaluation

The technique as described in the preceding section has been implemented using MATLAB (MathWorks, Inc., Natick, MA, USA) for single-pass processing of the input speech files. The implementation uses a total storage of 725 variables and coefficients (253 for envelope calculation in (7), 3 for dynamic range compression in (8), 1 for compressed signal in (9), 302 for Hilbert envelope in (10)-(12), 47 for smoothed peak in (13)-(14) and Figure 5, 5 for differentiation in (15), 103 for amplitude thresholding, and 11 for duration thresholding). The technique involves an algorithmic delay of 21.4 ms (12.5 ms for compression, 7.5 ms for Hilbert envelope, and 1.4 ms for epoch marking).

An example of the processing by the implementation of the proposed technique is shown in Figure 6 for the utterance /awa/ of a male speaker. Simultaneously recorded electro-glottogram (EGG) signal is also shown. The detected epochs are in accordance with the peaks of the glottal closure as seen in the negative of the differentiated EGG signal (DEGG).

A detailed performance evaluation of the technique was carried out using the CMU-ARCTIC database [28], with simultaneously recorded speech and EGG signals from five speakers and having recordings of 1132 sentences from two male and one female speaker, nonsense words from one male speaker, and 452 TIMIT sentences from one male speaker. For use in our testing, the speech and EGG recordings were down-sampled to 10 kHz and aligned using a delay adjustment of 0.7 ms [10]. The epochs detected using EGG were used as the reference epochs. Negative peaks of the first difference of the EGG signal represent the glottal closure instants [29] and these are marked using an adaptive thresholding. The RMS value of the entire DEGG record is used as the initial

Table 1: *Results of epoch detection on clean speech (mean and s.d. of the performance measures for 5 subjects).*

Method	IDR (%)		MR (%)		FR (%)		IDA (ms)		A-0.25 (%)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
SEDREAMS	93.9	3.2	5.6	2.5	0.4	0.7	0.6	0.2	76.7	22.7
ZFR	92.2	2.2	7.4	2.2	0.4	0.4	0.7	0.3	56.0	23.1
DYPSA	91.1	5.9	6.3	1.6	2.6	4.5	0.7	0.2	74.4	17.2
HEPD	**90.4**	**4.8**	**8.0**	**2.3**	**1.6**	**3.0**	**0.6**	**0.2**	**50.2**	**26.0**

Table 2: *Results of epoch detection on telephony speech (mean and s.d. of the performance measures for 5 subjects).*

Method	IDR (%)		MR (%)		FR (%)		IDA (ms)		A-0.25 (%)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
SEDREAMS	85.5	9.4	5.4	2.7	9.0	8.7	0.6	0.1	49.0	13.7
ZFR	63.9	29.2	5.9	3.3	30.3	28.9	0.6	0.2	51.1	20.5
DYPSA	89.9	6.5	7.2	2.8	3.0	4.0	0.5	0.1	81.2	10.9
HEPD	**87.5**	**6.8**	**10.0**	**2.6**	**2.5**	**5.0**	**0.5**	**0.2**	**68.2**	**16.2**

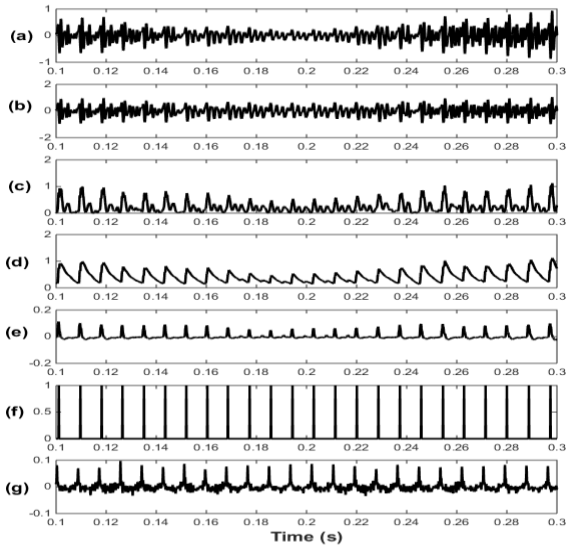

Figure 6: *Example of processing of the proposed HEPD technique: (a) input speech, (b) dynamic range compressed signal, (c) Hilbert envelope, (d) peak detector output, (e) differentiator output, (f) detected epochs (g) DEGG signal.*

threshold. When the amplitude of the DEGG signal exceeds the threshold, the sample corresponding to the highest point in the subsequent 1-ms interval is marked as the epoch and the threshold is updated as 0.3 times the amplitude of the epoch. Due to the insignificance of epochs in the unvoiced segments of speech, the evaluation is carried out only for the voiced segments, as detected using the DEGG signal. The interval between two instants of significant excitation is called as voiced if it corresponds to a pitch of 70 – 500 Hz.

Considering the larynx cycle as the interval between the successive epochs in the EGG signal, the following performance measures, as described in [20], are used to evaluate the performance of the technique:

- Identification rate (IDR): percentage of the larynx cycles with exactly one detected epoch.
- Miss rate (MR): percentage of the larynx cycles with no detected epoch.
- False alarm rate (FR): percentage of the larynx cycles with more than one detected epochs.
- Identification accuracy (IDA): standard deviation of the timing error between the reference and estimated epochs for the larynx cycles with one detected epoch.
- Accuracy to ±0.25 ms: percentage of the detected larynx cycles with the misalignment of the detected epoch with the reference epoch not exceeding 0.25 ms.

5. Test results

The proposed technique, based on the Hilbert envelope and peak detector, is referred to as HEPD for reporting the evaluation results. Its performance was compared with the ZFR, SEDREAMS, and DYPSA techniques, using the DYPSA and SEDREAMS implementations from [30] and [31], respectively. The performance was evaluated on the speech material of the CMU-ARCTIC database, with a total of 902718 epochs in the voiced segments. The performance measures were calculated for speech material from each of the five speakers separately and were used to find the mean and the standard deviation of the measures across the five speakers.

The results are given in Table 1 for the clean speech. In terms of the identification rate (IDR) and miss rate (MR), the SEDREAMS has the best performance. In terms of accuracy to ±0.25 ms (A-0.25), the performances of DYPSA and SEDREAMS are better than the other two techniques. Considering all performance measures, the techniques can be ranked as SEDREAMS, DYPSA, ZFR, and HEPD, with minor differences in their performances. Evaluation was also carried out on telephone-quality speech, which was simulated by bandpass filtering the speech signal with an approximate bandwidth of 300 – 3400 Hz, according to ITU-T P.862 [32]. These results are shown in Table 2. Considering all performance measures, the techniques may be ranked as DYPSA, HEPD, SEDREAMS, and ZFR, with ZFR generally giving poor results with lower mean and higher standard deviation. Thus, the two sets of results show that the performance of the proposed technique is close to the state-of-the-art techniques for clean speech and that it is robust against highpass filtering.

6. Conclusions

An epoch detection technique using Hilbert envelope and dynamic peak detection has been proposed. The method is well suited for real-time processing applications as it can be implemented with single-pass processing with an algorithmic delay of less than 30 ms and low memory requirements. The technique is validated using CMU-ARCTIC database and compared with state-of-the-art techniques and tested for robustness for telephone-quality speech. It needs to be further evaluated on larger speech databases and also for speech signals with pathologic voices.

Acknowledgments

The research is supported by 'National Program on Perception Engineering Phase-II,' sponsored by the Department of Electronics & Information Technology, Government of India.

References

[1] L. R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*, New Jersey: Prentice-Hall, 1975.

[2] J. L. Flanagan, *Speech Analysis Synthesis and Perception*, New York: Springer, 1972.

[3] B. Yegnanarayana, S. R. M. Prasanna, R. Duraiswami, and D. Zotkin, "Processing of reverberant speech for time-delay estimation," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 6, pp. 1110–1118, 2005.

[4] P. Lieberman, "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges," *J. Acoust. Soc. Amer.*, vol. 35, no. 3, pp. 344–353, 1963.

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

[6] J. P. Teixeira and A. Gonçalves, "Accuracy of jitter and shimmer measurements," in *Proc. Int. Conf. Health and Social Care Information Systems and Technologies (HCist 2016)*, Porto, Portugal, 2016, pp. 1190–1199.

[7] J. P. Teixeira and A. Gonçalves, "Algorithm for jitter and shimmer measurement in pathologic voices," in *Proc. Int. Conf. Health and Social Care Information Systems and Technologies (HCist 2016)*, Porto, Portugal, 2016, pp. 271–279.

[8] H. Kasuya, K. Masubuchi, S. Ebihara, and H. Yoshida, "Preliminary experiments on voice screening," *J. Phonetics*, vol. 14, no. 3, pp. 463–468, 1986.

[9] B. Yegnanarayana and R. N. J. Veldhuis, "Extraction of vocal-tract system characteristics from speech signals," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 313–327, 1998.

[10] 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, 2008.

[11] T. Drugman and T. Dutoit, "Glottal closure and opening instant detection from speech signals," in *Proc.* 10th *Annu. Conf. Int. Speech Commun. Assoc.* (*INTERSPEECH 2009)*, Brighton, UK, 2009, pp. 2891–2894.

[12] H. A. Patil and S. Viswanath, "Effectiveness of Teager energy operator for epoch detection from speech signals," *Int. J. Speech Technology*, vol. 14, no. 4, pp. 321–337, 2011.

[13] N. Shikhah and M. Deriche, "A novel pitch estimation technique using the Teager energy function," in *Proc. IEEE Int. Symp. Signal Processing and Applications (ISSPA 1999)*, Brisbane, Queensland, Australia, 1999, pp. 135–138.

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

[15] 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 Sig. Process. Letters*, vol. 14, no. 10, pp. 762–765, 2007.

[16] 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, Lang. Process.*, vol. 20, no. 3, pp. 994–1006, 2012.

[17] 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, Lang. Process.*, vol. 21, no. 12, pp. 2471–2480, 2013.

[18] A. P. Prathosh, P. Sujith, A. G. Ramakrishnan, and P. K. Ghosh, "Cumulative impulse strength for epoch extraction," *IEEE Sig. Process. Letters*, vol. 23, no. 4, pp. 424 –428, 2016.

[19] 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, 1995.

[20] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes, "Estimation of glottal closure instants in voiced speech using the DYPSA algorithm," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 34–43, 2007.

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

[22] K. Vijayan and K. S. R. Murty, "Epoch extraction by phase modeling of speech signals," *Circuits, Syst., Signal Process.*, vol. 35, no. 7, pp. 2584–2609, 2016.

[23] 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. ASSP-27, no. 4, pp. 309–319, 1979.

[24] International Telecommunication Union: Relative timing of sound and vision for broadcasting, ITU Rec. ITU-R BT.1359 1998.

[25] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, Upper Saddle River, New Jersey: Prentice-Hall, 1999.

[26] U. Zölzer, *Digital Audio Signal Processing*, West Sussex, United Kingdom: John Wiley & Sons, 2008.

[27] L. R. Rabiner, M. Sambur, and C. E. Schmidt, "Applications of a nonlinear smoothing algorithm to speech processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 23, no. 6, pp. 552–557, 1975.

[28] *The Festvox Website*, [Online]. Available: festvox.org

[29] D. G. Childers and A. K. Krishnamurthy, "A critical review of electroglottography," *CRC Crit. Rev. Boeing.*, vol. 12, pp. 131–164, 1985.

[30] M. Brookes, *Voicebox*, [Online].Available: www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[31] T. Drugman, *Gloat toolbox*, [Online]. Available: tcts.fpms.ac.be/~drugman/Toolbox/.

[32] P. Loizou, and Y.Hu, *NOIZEUS: A noisy speech corpus for evaluation of speech enhancement algorithms*, [Online]. Available:ecs.utdallas.edu/loizou/speech/noizeus/.