EPOCH EXTRACTION FROM ALLPASS RESIDUAL OF SPEECH SIGNALS

Karthika Vijayan and K. Sri Rama Murty

Department of Electrical Engineering, Indian Institute of Technology Hyderabad, India.

ee11p011@iith.ac.in, ksrm@iith.ac.in

ABSTRACT

Identification of epochs from speech signals is a prominent task in speech processing. In this paper, epoch extraction is attempted from phase spectrum of speech signals. The phase spectrum of speech is modelled as an allpass (AP) filter by minimizing entropy of energy in the associated error signal. The AP residual thus obtained contains prominent unambiguous peaks at epoch locations. These peaks in AP residual constitute a set of candidate epoch locations from which appropriate ones are identified using a dynamic programming algorithm. The proposed method is evaluated on a subset of CMU Arctic database and it is observed that it delivered better epoch extraction performance than the prominent speech events estimation method-DYPSA. In case of telephone channel speech, the proposed method significantly outperformed zero frequency resonator based method also.

Index Terms— Epochs, Phase spectrum, Allpass modelling, Entropy, Allpass residual.

1. INTRODUCTION

Voiced sounds are produced by exciting the vocal-tract system (VTS) with quasi-periodic sequence of glottal pulses, generated by vibration of vocal-folds. Within each period of a glottal pulse, significant excitation of VTS happens at the instant of glottal-closure, which is called the *epoch* [1]. Many speech analysis methods benefit from accurate estimation of epochs from speech signal. For example, knowledge of epoch locations can be used to estimate instantaneous fundamental frequency from speech signals [2]. The region immediately after an epoch is referred to as glottal closure region, whose properties were exploited in computing free resonances of VTS [3]. The information about glottal closure regions was employed for speech enhancement, as they are less affected by noise [4]. The knowledge of epochs is used for pitch synchronous analysis of speech and prosody modification [5, 6].

Speech is the outcome of a time-varying VTS driven by a time-varying excitation source. The information about epoch locations is mainly manifested in the excitation source of speech signal. The VTS characteristics needs to be suppressed to highlight information about excitation source. This can be achieved by modelling VTS, and performing inverse filtering of speech through the model. Linear prediction (LP) analysis is one such method to model VTS as an all-pole filter [7]. The filter coefficients are used to inverse filter speech signal to derive LP residual. Since LP analysis models only the magnitude response of VTS, LP residual contains multiple peaks of either polarity around epoch locations, as it can be seen from Fig. 1(b). As a result, epoch locations cannot be identified unambiguously from LP residual. In order to address this issue, Ananthapadmanabha et al. used Hilbert envelope of LP residual to identify epoch locations [8]. Smits et al. used phase slope function derived from the group-delay of LP residual for epoch extraction [9]. Naylor et al. proposed a dynamic programming based algorithm (DYPSA) for epoch extraction using phase slope functions and quasi-periodic nature of epoch locations [10, 11]. Drugman et al. presented a mean-based signal and LP residual based method (SEDREAMS) for epoch identification together with a review of several methods in [12]. In all the above methods, VTS characteristics are suppressed using the LP analysis. In zero-frequency resonator (ZFR) method, the speech signal is passed through a narrowband infinite-impulse response (IIR) filter around 0 Hz to suppress the characteristics of VTS [13]. Since vocal-tract resonances exist only above 300 Hz, the output of ZFR mainly contains the information about excitation source. The zero-crossings of the output of ZFR are hypothesized as epoch locations. An FIR implementation of ZFR based method along with dynamic programming was studied in [14].

The information about epoch locations is mainly reflected in the phase spectrum of speech signal. In this paper, we propose a method for epoch extraction from the phase spectrum of speech. The phase spectrum of a segment of speech signal is modelled using an allpass (AP) filter. The error signal associated with the modelling, referred to as AP residual, contains significant unambiguous peaks at epoch locations. The peaks in AP residual form the candidates for epoch locations from which suitable ones are selected using a dynamic programming algorithm.

The rest of the paper is organized as follows: Section 2 discusses our motivation for modelling phase spectrum of speech signals. Section 3 describes a method for estimation of AP filter coefficients and computation of AP residual. Experimental evaluation of the proposed method and its comparison with existing methods is presented in Section 4. In section 5 we conclude the proposed work.

2. ALLPASS MODELLING OF PHASE SPECTRUM

The frequency domain representation $S(j\omega)$ of a discrete-time signal $s[n]$ is given by [15]

$$S(j\omega) = \sum_{n=-\infty}^{+\infty} s[n]e^{-j\omega n} \tag{1}$$

which can be expressed in the polar form as

$$S(j\omega) = |S(j\omega)|e^{j\angle S(j\omega)} \tag{2}$$

where $|S(j\omega)|$ is called the magnitude spectrum and $\angle S(j\omega)$ is called the phase spectrum of signal $s[n]$. In the case of speech signals, the magnitude spectral envelope is mainly contributed by VTS,

The authors would like to acknowledge the Department of Electronics and Information Technology (DeitY), Ministry of Communications & Information Technology, Government of India for sponsoring this work.

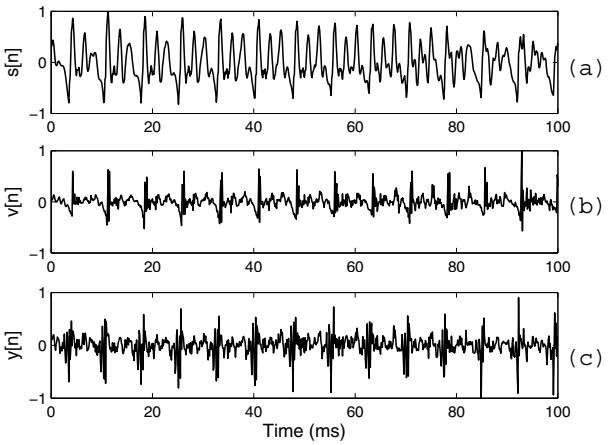

Fig. 1: Demonstration of characteristics of phase signal: (a) Speech signal, (b) LP residual and (c) Phase signal.

while the phase spectrum is mainly contributed by excitation source. Hence the information about VTS can be suppressed by removing the magnitude spectral information from speech signal. Magnitude spectral information can be removed by dividing the speech spectrum $S(j\omega)$ with its magnitude spectrum $|S(j\omega)|$. A signal, termed as phase signal, which retains only the phase spectral characteristics of $s[n]$ is derived as follows

$$y[n] = \mathcal{F}^{-1}\left\{\frac{S(j\omega)}{|S(j\omega)|}\right\} \tag{3}$$

where $\mathcal{F}^{-1}$ denotes the inverse discrete-time Fourier transform. Removing magnitude spectral information in frequency-domain is equivalent to removing second order relations among the samples in time-domain. The samples of phase signal $y[n]$ are uncorrelated, owing to the magnitude spectrum removal, but they are not independent. That is, the information about phase spectrum of $s[n]$ is embedded in the higher order relations among samples of $y[n]$. The phase signal $y[n]$ derived from a speech segment $s[n]$ in Fig. 1(a) is shown in Fig. 1(c). The phase signal $y[n]$ in Fig. 1(c) contains multiple peaks of either polarity around epoch locations due to the phase spectral characteristics induced by VTS. These phase spectral characteristics have to be suppressed in order to enhance the evidence for true epoch locations.

An investigation for a suitable filter for modelling phase signal $y[n]$ leads to allpass (AP) filter. AP filter is an autoregressive moving average (ARMA) filter with zeros of the transfer function lying at conjugate reciprocal locations of its poles. The transfer function of an M^{th} order AP filter is given by

$$H(z) = \frac{a_M + a_{M-1}z^{-1} + ... + a_1 z^{-M+1} + z^{-M}}{1 + a_1 z^{-1} + ... + a_{M-1}z^{-M+1} + a_M z^{-M}} \tag{4}$$

Both the numerator and denominator polynomials of $H(z)$ have same set of coefficients $\mathbf{a} = [a_1 a_2 ... a_M]^T$, one being the flipped version of other. It can be shown that the AP filter in (4) exhibits a flat magnitude response ($|H(j\omega)| = 1$), and an arbitrary phase response depending on the filter coefficients $\mathbf{a}$. An AP filter, when excited with a sequence of independent and identically distributed (i.i.d.) non-Gaussian samples $x[n]$, generates a sequence of uncorrelated but dependent samples as output. This characteristic exactly matches with that of the phase signal $y[n]$, and hence AP filter is a suitable choice for modelling $y[n]$.

In this work, we have modelled the phase signal $y[n]$ as the output of an AP filter excited with i.i.d. non-Gaussian input $x[n]$. Given the output signal $y[n]$ and the filter coefficients $\mathbf{a}$, the input signal $x[n]$ can be computed in a noncausal manner as follows:

$$x[n] = -\sum_{l=1}^{M} a_l x[n+l] + y[n+M] + \sum_{l=1}^{M} a_l y[n+M-l] \tag{5}$$

The coefficients of the AP filter are estimated by imposing production-specific constraints on the input $x[n]$.

3. ESTIMATION OF ALLPASS COEFFICIENTS

Estimation of AP transfer function is an ill-posed problem, because both filter coefficients $\mathbf{a}$ and input $x[n]$ are unknown. Hence it requires some assumptions on either $\mathbf{a}$ or $x[n]$ to deal with the modelling problem. Chi et al., proposed a method for estimating the filter coefficients $\mathbf{a}$ which maximizes the higher-order cumulant (3^{rd} or 4^{th}) of the input signal $x[n]$ [16]. The AP modelling of LP residual using this method was presented in [17] with application to speaker recognition. This method requires the knowledge of optimal cumulant to maximize. Breidt et al. modelled financial time series as an output of AP filter by enforcing Laplacian distribution on input signal $x[n]$ [18]. A maximum likelihood estimation of AP systems is proposed in [19] where the input signal $x[n]$ follows an arbitrary probability distribution with known parameters. These methods require apriori information about the probability distribution followed by the input signal $x[n]$. In this paper, we estimate the AP filter coefficients by imposing speech production specific constraints on the input signal $x[n]$.

From the acoustic theory of speech production, it is known that significant component of the energy in each glottal cycle is delivered at the instant of glottal closure. An ideal scenario can be visualized as an impulse excitation at the epoch, and zero excitation elsewhere in the glottal cycle. The input to the voiced segment can be assumed as a sequence of impulses, in which the energy is concentrated only at a few samples. We would like to derive an input signal $x[n]$ whose energy is concentrated only at epoch locations. Since $x[n]$ is derived by inverse filtering the phase signal $y[n]$ through the AP filter, the total energy of $x[n]$ should be same as the total energy of $y[n]$. So without loss of generality, we can make the phase signal $y[n]$ a unit energy signal by dividing each sample with the total energy of the frame. We need to estimate the AP filter coefficients such that the unit energy in $x[n]$ gets concentrated only at a few samples. The energy contributed by each sample in the input signal $x[n]$ can be expressed as

$$e[n] = x^2[n] \tag{6}$$

Since each of the samples in $e[n]$ is positive and sum of the samples is unity, it can be viewed as a valid probability mass function. The energy $e[n]$ can be made to concentrate only at a few samples by minimizing its entropy. The concept of minimum entropy deconvolution was explored for seismic recordings in [20], which was later extended to period estimation of quasi-periodic sequences in [21]. In this paper, we adopt a gradient based iterative procedure for entropy minimization. The entropy of $e[n]$ can be expressed as a function of AP filter coefficients $\mathbf{a}$ as [22]:

$$J(\mathbf{a}) = -\sum_{n=1}^{N} e[n]\ log\ e[n] \tag{7}$$

The AP coefficients $\mathbf{a}$ are estimated by minimizing the entropy in (7). In the minimization procedure, the set of filter coefficients $\mathbf{a}_k$ at iteration k is computed as

$$\mathbf{a}_k = \mathbf{a}_{k-1} - \eta \frac{\partial J(\mathbf{a})}{\partial \mathbf{a}} \Big|_{\mathbf{a}=\mathbf{a}_{k-1}} \quad (8)$$

where $\mathbf{a}_{k-1}$ are the filter coefficients at $(k-1)^{th}$ iteration, η is the learning rate parameter, and $\frac{\partial J(\mathbf{a})}{\partial \mathbf{a}}$ is the gradient. At the beginning of the iterations, $\mathbf{a}$ is initialized to small random values ensuring that the poles of $H(z)$ are lying inside unit circle. The gradient of the objective function $J(\mathbf{a})$ with respect to the filter coefficients $\mathbf{a}$ is calculated using chain rule as follows:

$$\begin{aligned} \frac{\partial J(\mathbf{a})}{\partial \mathbf{a}} &= \frac{\partial J(\mathbf{a})}{\partial e[n]} \frac{\partial e[n]}{\partial x[n]} \frac{\partial x[n]}{\partial \mathbf{a}} \\ &= -\sum_{n=1}^{N} (1 + \log e[n]) (2x[n]) \left(\frac{\partial x[n]}{\partial \mathbf{a}} \right) \end{aligned} \quad (9)$$

where, the derivative of $x[n]$ with respect to $\mathbf{a}$ is given by

$$\frac{\partial x[n]}{\partial a_p} = -\sum_{l=1}^{M} a_l \frac{\partial x[n+l]}{\partial a_p} - x[n+p] + y[n+M-p] \quad (10)$$

for $\forall p \in \{1, 2...M\}$. A careful observation of (10) suggests that the derivative of $x[n]$ with respect to a_p can be computed by filtering $y[n+M-p] - x[n+p]$ through an all-pole filter with coefficients $\mathbf{a}$. The filter coefficients are updated using the above equations till the decrement in J becomes negligibly small. The updating of coefficients will be terminated when the decrement in J is smaller than a predefined threshold, i.e. $J(\mathbf{a}_{k-1}) - J(\mathbf{a}_k) < \epsilon$, where threshold value- ϵ is chosen as 10^{-6} in this study.

The behaviour of objective function for voiced and unvoiced segments is shown in Fig. 2. The objective function achieved a lower value for voiced segment than for an unvoiced segment. This is because the excitation for voiced segments is impulse-like and is concentrated only at the epoch locations. On the other hand, the excitation for unvoiced segments is noise-like and spreads across the time which cannot be modelled effectively.

3.1. Epoch extraction

The estimated AP coefficients are used to inverse filter the phase signal $y[n]$ to obtain an estimate of the input signal $x[n]$, which hereafter, is referred to as *AP residual*. The AP residual obtained from a voiced speech segment exhibits sharper peaks at the epoch locations as entropy of its energy is minimized. Fig. 3 shows absolute values of AP residual obtained for the speech segment shown in Fig. 1(a) by modelling its phase signal using AP filters of different orders. The 14^{th} order AP residual in Fig. 3(b) has unambiguous peaks at the epoch locations which coincide exactly with the ground truth from differenced electroglottograph (DEGG) in Fig. 3(d). If the model order is too low, then resulting AP residual contains multiple peaks around epoch locations as shown in Fig. 3(a). On the other hand, a higher order model might overfit the phase spectrum, leading to failure in detecting some epoch locations as shown in Fig. 3(c) (60-80 ms). Notice that the total energy of the segments in Fig. 3(a), Fig. 3(b) and Fig. 3(c) are exactly same, but the distribution of energy differs depending on the model order. It is observed that a model order between 10-15 suits well for epoch extraction.

The group-delay spectrum of the estimated AP filter coefficients is shown in Fig. 4(a) to interpret the information captured by the model. The group-delay spectrum in Fig 4(a) shows clear peaks at the resonant frequencies of VTS, and matches exactly with those in the LP spectrum given in Fig. 4(b) for reference. It is interesting to note that the information about the resonances is still preserved in the phase signal $y[n]$ even though its magnitude spectrum is flat.

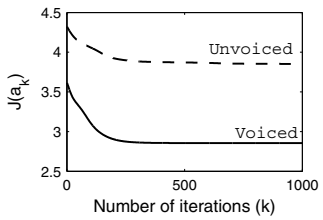

Fig. 2: Behaviour of the objective function for voiced and unvoiced frames.

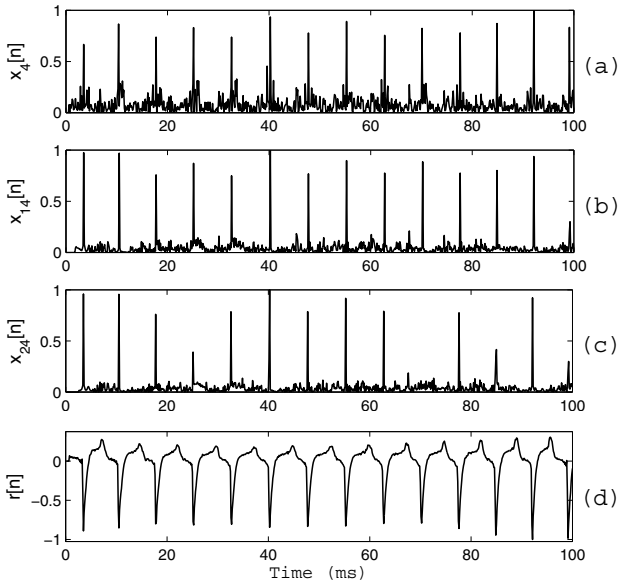

Fig. 3: Effect of model order M on AP residual. Absolute values of AP residual obtained with (a) $M = 4$, (b) $M = 14$, (c) $M = 24$ and (d) DEGG signal for reference.

3.2. Epoch selection

The AP residual contains unambiguous peaks of single sample width at epoch locations and relatively small values elsewhere. This characteristic of AP residual is exploited to identify the candidates for epoch locations. From these candidate locations, epochs are selected by minimizing a cost function using dynamic programming approach similar to the one proposed in [10]. The value of a cost function (penalty) is minimized using a P-best dynamic programming algorithm [23, 24]. The factors used in constructing the cost function with respect to a matrix of candidate epoch locations are: strength of the peak in AP residual, correlation coefficient between frames of speech, and Frobenius norm cost [25]. Unlike general dynamic programming approaches which deliver single path with minimum penalty, the P-best algorithm will keep P paths with least penalties at each iteration. These paths will act as history for the next iteration in deciding the penalty values of intersecting paths at differ-

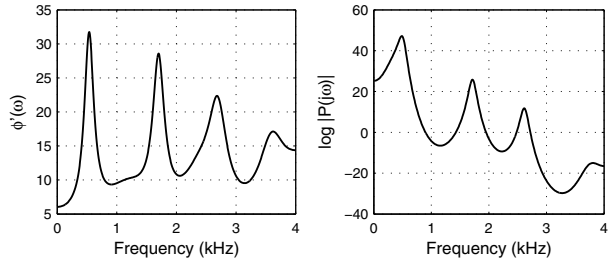

Fig. 4: (a) Group-delay spectrum obtained from estimated AP coefficients, (b) LP spectrum for reference

ent nodes in the matrix of candidate epoch locations. This algorithm gives robust and unambiguous decisions of right epoch locations by considering a beam of search space rather than following single absolute path in the epoch locations matrix [23].

4. EXPERIMENTAL EVALUATION

A subset of CMU Arctic database consisting of 100 utterances each from a male (bdk) and a female speaker (slt) is used to evaluate the proposed method. The electroglottograph (EGG) signals are used to extract the reference epoch locations. The database consists of 52125 epoch locations. All the signals are down-sampled to 8 kHz for evaluation.

Speech signal is windowed into frames of 25 ms shifted by 10 ms. For each frame of 25 ms, the magnitude spectrum is suppressed to obtain the phase signal $y[n]$ as in (3). The phase signal $y[n]$ is modelled as the output of AP filter, and the AP residual $x[n]$ is derived. In this study we have used a 14^{th} order AP filter to model the phase signal $y[n]$. AP residual $x[n]$ is thresholded to 0.2 to obtain the candidates for epoch locations. Selection of genuine epochs from the candidates was done using a P-best dynamic programming approach, where P is chosen as 5. The selected epochs are compared with the reference epoch locations to evaluate the performance.

The performance evaluation is carried out by defining the following measures [10]:

1. Identification rate (IDR) is defined as the number of larynx cycles in which exactly one epoch is detected
2. Miss rate (MR) is the number of larynx cycles where no epochs has been detected
3. False alarm rate (FAR) is the number of larynx cycles where more than one epoch locations have been returned.
4. Identification error (ζ) is the timing error between the reference epoch location and the detected epoch location in case of proper identification.

where a larynx cycle is defined as $\frac{1}{2}(l_{r-1}+l_r) \leq n \leq \frac{1}{2}(l_r+l_{r+1})$, where l_r is a reference epoch location with preceding and succeeding reference epoch locations at l_{r-1} and l_{r+1}.

The performance of the proposed method based on AP modelling (APM), along with two other prominent epoch extraction methods namely ZFR [13] and DYPSA [10] (implemented using [26]), is given in Table 1. The proposed method performs better than the LP residual based DYPSA, although the performance of ZFR method is the best among the three.

Evaluation of the proposed method is done on telephone speech to check for robustness towards telephone channel effects. Telephone channel is simulated as a bandpass infinite impulse response

Table 1: Performance of epoch extraction methods on clean speech.

Method	IDR (%)	MR (%)	FAR (%)	ζ (ms)
APM	96.42	2.24	1.34	0.3442
ZFR	98.03	0.62	1.35	0.3195
DYPSA	91.71	7.25	1.04	0.3586

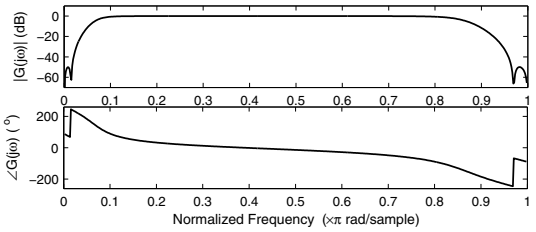

Fig. 5: Frequency response of the telephone channel filter (for sampling frequency= 8kHz).

(IIR) filter, $G(z)$ for the channel bandwidth of 300 Hz to 3400 Hz whose frequency response is shown in Fig. 5. Clean speech from the database is filtered out through this filter to obtain telephone speech. The performances of different epoch extraction strategies for telephone speech is shown in Table. 2. The epoch identification rate of the proposed method for telephone speech stayed relatively equivalent to that for clean speech. DYPSA stayed inferior to the proposed algorithm, though it delivered a slightly better performance for telephone speech. But the performance of ZFR drastically reduced since it relies on the impulse-like nature of excitation which was brought out by filtering around zero frequency. The telephone speech is deprived of components around zero frequency and hence ZFR fails to capture excitation information.

Table 2: Performance of epoch extraction methods on telephone speech.

Method	IDR (%)	MR (%)	FAR (%)	ζ (ms)
APM	93.68	3.55	2.77	0.3521
ZFR	70.06	2.08	27.86	0.3895
DYPSA	92.26	6.56	1.18	0.3606

5. CONCLUSIONS

An epoch extraction methodology relying on phase spectral characteristics of speech signals was proposed in this paper. Phase spectrum of speech was modelled as an allpass (AP) filter by minimizing the entropy of energy content in AP residual. The resultant AP residual exhibited train of impulse-like nature with proper placement of prominent impulses at epoch locations, thus rendering unambiguous epoch identification. The epoch extraction performance of the proposed method was found to be superior to DYPSA algorithm, and in case of telephone speech, it outperformed zero frequency based method too. These results pointed to the evidence for presence of information about epochs in phase spectrum of speech, which was captured by AP modelling and was manifested in AP residual as sharp peaks.

6. REFERENCES

[1] Jan Gauffin and Johan Sundberg, "Spectral correlates of glottal voice source waveform characteristics," *Journal of Speech, Language and Hearing Research*, vol. 32, no. 3, pp. 556, 1989.

[2] B. Yegnanarayana and K. Sri Rama Murty, "Event-based instantaneous fundamental frequency estimation from speech signals," in *IEEE Transactions on Audio, Speech and Language Processing*, May 2009, vol. 17, pp. 614–624.

[3] B. Yegnanarayana and Raymond N. J. Veldhuis, "Extraction of vocal-tract system characteristics from speech signals," in *IEEE Transactions on Speech and Audio Processing*, July 1998, vol. 6, pp. 313–327.

[4] B. Yegnanarayana and P. S. Murty, "Enhancement of reverberent speech using lp residual signal," in *IEEE Transactions on Speech and Audio Processing*, May 2000, vol. 8, pp. 267–281.

[5] K. Sreenivasa Rao and B. Yegnanarayana, "Prosody modification using instants of significant excitation," in *IEEE Transactions on Audio, Speech and Language Processing*, May 2006, vol. 14, pp. 972–980.

[6] K. Sreenivasa Rao and B. Yegnanarayana, "Duration modification using glottal closure instants and vowel onset points," in *Speech Communication*, 2009, vol. 51, pp. 1263–1269.

[7] John Makhoul, "Linear prediction: A tutorial review," in *Proceedings of the IEEE*, April 1975, vol. 63, pp. 561–580.

[8] T. V. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction from linear prediction residual for identification of closed glottis interval," in *IEEE Transactions on Acoustics, Speech and Signal Processing*, August 1979, vol. ASSP-27, pp. 309–319.

[9] Roel Smits and B. Yegnanarayana, "Determination of instants of significant excitation in speech using group delay function," in *IEEE Transactions on Speech and Audio Processing*, September 1995, vol. 3, pp. 325–333.

[10] Patrick A. Naylor, A. Kounoudes, Jon Gudnason, and Mike Brookes, "Estimation of glottal closure instants in voiced speech using the dypsa algorithm," in *IEEE Transactions on Audio, Speech and Language Processing*, January 2007, vol. 15, pp. 34–43.

[11] A. Kounoudes, Patrick A. Naylor, and Mike Brookes, "The dypsa algorithm for estimation of glottal closure instants in voiced speech," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2002, pp. 349–352.

[12] Thomas Drugman, Mark Thomas, Jon Gudnason, Patrick Naylor, and Thierry Dutoit, "Detection of glottal closure instants from speech signals: A quantitative review," in *IEEE Transactions on Audio, Speech and Language Processing*, March 2012, vol. 20, pp. 994–1006.

[13] K. Sri Rama Murty and B. Yegnanarayana, "Epoch extraction from speech signals," in *IEEE Transactions on Audio, Speech and Language Processing*, November 2008, vol. 16, pp. 1602–1613.

[14] Mark R. P. Thomas, Jon Gudnason, and Patrick A. Naylor, "Estimation of glottal closing and opening instants in voiced speech using the yaga algorithm," in *IEEE Transactions on Audio, Speech and Language Processing*, January 2012, vol. 20, pp. 82–91.

[15] Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck, *Discrete-time signal processing*, Signal processing series. Prentice Hall, Inc., Upper Saddle River, NJ, USA., 2nd edition, January 1999.

[16] C.-Y. Chi and J.-Y. Kung, "A new identification algorithm for allpass systems by higher-order statistics," in *Signal Processing*, January 1995, vol. 41, pp. 239–256.

[17] K. Sri Rama Murty, Vivek Boominathan, and Karthika Vijayan, "Allpass modeling of lp residual for speaker recognition," in *International Conference on Signal Processing and Communications, SPCOM*, July 2012, pp. 1–5.

[18] F. J. Breidt, R. A. Davis, and A. A. Trindade, "Least absolute deviation estimation for all-pass time series models," in *Annals of statistics*, 2001, vol. 29, pp. 919–946.

[19] B. Andrews, R. A. Davis, and F. J. Breidt, "Maximum likelihood estimation for all-pass time series models," in *Journal of Multivariate Analysis*, August 2006, vol. 97, pp. 1638–1659.

[20] Ralph A. Wiggins, "Minimum entropy deconvolution," in *Geoexploration*, April 1978, vol. 16, pp. 21–35.

[21] G. González, R.E. Badra, R. Medina, and J. Regidor, "Period estimation using minimum entropy deconvolution (med)," in *Signal Processing*, January 1995, vol. 41, pp. 91–100.

[22] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Telecommunications and signal processing. Wiley-Interscience, 2006.

[23] Richard Schwartz and Yen-Lu Chow, "The n-best algorithm: an efficient and exact procedure for finding the n most likely sentence hypotheses," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1990, pp. 81–84.

[24] Jung-Kuei Chen and Frank K. Soong, "An n-best candidates-based discriminative training for speech recognition applications," in *IEEE Transactions on Speech and Audio Processing*, January 1994, vol. 2, pp. 206–216.

[25] Changxue Ma, Yves Kamp, and Lei F. Willems, "A frobenius norm approach to glottal closure detection from the speech signal," in *IEEE Transactions on Speech and Audio Processing*, April 1994, vol. 2, pp. 258–265.

[26] Mike Brookes, "VOICEBOX: Speech Processing Toolbox for MATLAB," Available Online, 2005.