



Prosody Modification using Allpass Residual of Speech Signals

Karthika Vijayan and K. Sri Rama Murty

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

{ee11p011, ksrm}@iith.ac.in

Abstract

In this paper, we attempt to signify the role of phase spectrum of speech signals in acquiring an accurate estimate of excitation source for prosody modification. The phase spectrum is parametrically modeled as the response of an allpass (AP) filter, and the filter coefficients are estimated by considering the linear prediction (LP) residual as the output of the AP filter. The resultant residual signal, namely AP residual, exhibits unambiguous peaks corresponding to epochs, which are chosen as pitch markers for prosody modification. This strategy efficiently removes ambiguities associated with pitch marking, required for pitch synchronous overlap-add (PSOLA) method. The prosody modification using AP residual is advantageous than time domain PSOLA (TD-PSOLA) using speech signals, as it offers fewer distortions due to its flat magnitude spectrum. Windowing centered around unambiguous peaks in AP residual is used for segmentation, followed by pitch/duration modification of AP residual by mapping of pitch markers. The modified speech signal is obtained from modified AP residual using synthesis filters. The mean opinion scores are used for performance evaluation of the proposed method, and it is observed that the AP residual-based method delivers equivalent performance as that of LP residual-based method using epochs, and better performance than the linear prediction PSOLA (LP-PSOLA).

1. Introduction

Prosody modification refers to the controlled alteration of loudness, pitch and duration of speech units [1]. It finds applications in concatenative speech synthesis, where a sequence of speech units is played in continuum and without perceivable distortions at unit boundaries [2, 3]. Duration expansion and compression are used in playback systems, for slowing down speech for better intelligibility and fast scanning of records, respectively [4]. Pitch modification is used in text-to-speech (TTS) systems and for voice conversion applications [2, 5].

The most commonly used method for duration and pitch modifications of speech signals is the pitch synchronous overlap and add (PSOLA) technique [3]. The time domain PSOLA (TD-PSOLA) technique modifies segments of speech obtained from pitch-synchronous windowing, and overlaps and adds the modified segments. But, the TD-PSOLA technique is affected with pitch, phase and spectral mismatches caused by inaccurate placement of windows [6]. The quality of synthesized speech from TD-PSOLA technique largely depends on the accuracy of pitch markers, upon which the windows are centered. The estimation and manual correction of pitch markers required by PSOLA, are tedious and expensive tasks. A variety of pitch marking algorithms were proposed for operation of PSOLA technique. The points of waveform similarity were identified as pitch markers using signal autocorrelation coef-

ficients, spectral correlation functions, absolute difference between successive segments of speech at different time lags, etc. [7, 8, 9, 10]. The pitch markers were also identified as points of highest short-time energy in speech signals [11, 12, 13]. The raw pitch markers identified using these techniques were refined using dynamic programming algorithms with similarity cost functions [9, 11] or by relying on the continuity of pitch contour [11, 10, 14]. The disadvantage of many of these methods is that, they bank on several ad-hoc parameters.

The PSOLA technique was also performed on the linear prediction (LP) residual [3]. As the LP residual has nearly flat magnitude spectrum, lesser spectral mismatches are incurred than PSOLA on speech signals. Like TD-PSOLA, accurate pitch markers are required by LP-PSOLA and the instants of significant excitation (epochs) are mostly used as pitch markers. Peak-picking from the speech signal [15], from Hilbert envelope [16] and using average group delay [17] were done for epoch extraction. The performance of LP-PSOLA largely depends on the accuracy of epoch extraction algorithms. The harmonic plus noise model, sinusoidal model, etc. and phase vocoders were also used for prosody manipulation [18, 1, 19].

In this paper, we propose to model the phase and magnitude spectra of speech signals to estimate a potentially complete model for vocal tract system (VTS) and an accurate representative of the excitation signal, for prosody modification. The magnitude spectrum is modeled using LP analysis and resultant LP residual is obtained. The phase spectrum is modeled by considering the LP residual as the response of an allpass (AP) filter. As the phase spectrum is modeled as an AP filter response, it does not manipulate the magnitude spectrum of signals. The resultant AP residual is a true representative of excitation source, and exhibits nearly flat magnitude spectrum like the LP residual, resulting in fewer spectral distortions in prosody modification. Also the AP residual holds unambiguous peaks at epochs, which are used as pitch markers for prosody modification, thereby removing ambiguities regarding placement of windows. Thus the AP residual encompasses the advantage with LP residual (nearly flat magnitude spectrum) and nullifies its disadvantage (ambiguities with pitch marking) for prosody modification. The short-time segmentation of AP residual is done by windowing around at its peaks, followed by pitch/duration modifications by altering the epochal pitch markers. The speech signals are reconstructed from modified AP residual using synthesis filters. Subjective analysis conducted to evaluate prosody modification shows the efficiency of the proposed method.

The rest of this paper is organized as follows: Section 2 describes the AP modeling strategy for phase spectrum of speech signals. In Section 3, the strategy for prosody modification using AP residual is elaborated. Section 4 discusses the results of subjective evaluation. Section 5 summarizes the contributions of this paper towards prosody modification.

2. Allpass modeling of phase spectrum

The discrete-time Fourier transform of a signal $s[n]$ is [20]:

$$\begin{aligned} S(j\omega) &= \sum_{n=-\infty}^{\infty} s[n]e^{-j\omega n} \\ &= |S(j\omega)|e^{j\angle S(j\omega)} \end{aligned} \quad (1)$$

where $|S(j\omega)|$ and $\angle S(j\omega)$ are the magnitude and phase spectra of $s[n]$. In this paper, we intend to obtain parametric models for both magnitude and phase spectra of speech signals, in order to realize a potentially complete model for VTS and to implement synthesis filters for speech reconstruction.

We perform LP analysis on short-time segments of speech signals, in order to model the envelope of magnitude spectrum and obtain the LP residual. The LP filter $G(z)$ models the VTS as a minimum phase all-pole filter, given by [21]:

$$G(z) = \frac{1}{1 + \sum_{k=1}^M a_k z^{-k}} \quad (2)$$

where M is the order of LP analysis and $\mathbf{a} = [a_1 a_2 \dots a_M]^T$ is the set of LP coefficients (LPCs). The LPCs are estimated by minimizing the mean square error between the true value of a sample and its predicted value (linear combination of past samples). The estimated LP filter approximates the gross spectral envelope of speech signals [21]. The magnitude spectrum and the modeled LP spectrum of a short-time segment of speech signal are shown in Figure 1(a) and Figure 1(b), respectively. The LP spectrum grossly coincides with the envelope of magnitude spectrum of speech signal, revealing information about resonances of VTS as observable peaks. The resultant LP residual can be obtained by inverse filtering speech signal through the estimated LP filter $G(z)$. A segment of speech signal $s[n]$ and resultant LP residual $y[n]$, are shown in Figure 2(a) and Figure 2(b), respectively. The LP residual exhibits multiple peaks of either polarity around epochs, due to the presence of unmodeled phase spectrum of speech signals.

For modeling the phase spectrum, the magnitude spectrum has to be removed to highlight the phase spectral characteristics. The LP residual can be viewed as a signal with suppressed magnitude spectrum, as it is generated as error in LP analysis, which models the magnitude spectrum. The LP residual has nearly flat magnitude spectral envelope as shown in Figure 1(c). Consequently, the samples of LP residual are marginally correlated. But it holds the unmodeled phase spectral information in speech after LP analysis, and hence has higher order statistical relationships between its samples [22]. We need to model the LP residual with nearly-uncorrelated, but dependent samples.

The AP filter generates uncorrelated and dependent output samples when excited with an independent and identically distributed (i.i.d.) input sequence $x[n]$. This characteristic makes the AP filter an appropriate choice for modeling the LP residual. The transfer function of the AP filter is given by [20]

$$H(z) = \frac{w_M + w_{M-1}z^{-1} + \dots + w_1z^{-M+1} + z^{-M}}{1 + w_1z^{-1} + \dots + w_{M-1}z^{-M+1} + w_Mz^{-M}} \quad (3)$$

The poles and zeros of $H(z)$ lie at conjugate reciprocal locations of each other, and hence the magnitude response of an AP filter is unity ($|H(j\omega)| = 1$). Thus the AP filter does not modify the magnitude spectrum of its input signal, consequently the energy of its input and output signals are the same. The transfer function of AP filter $H(z)$ is completely characterized by the

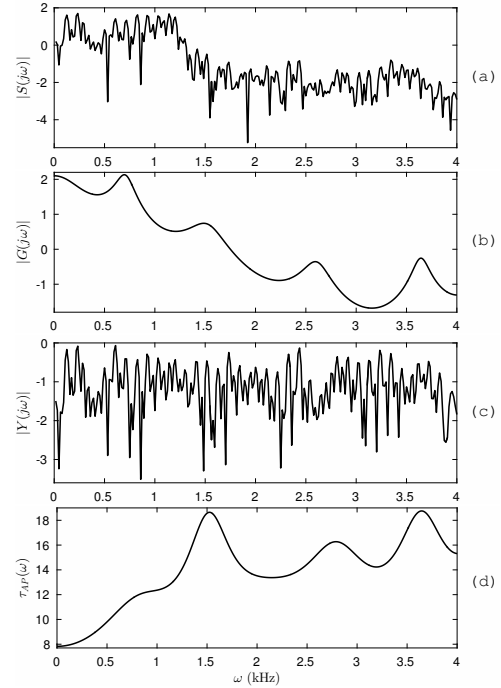


Figure 1: Illustration of efficacy of modeling strategies: (a) magnitude spectrum of speech (b) LP magnitude spectrum (c) magnitude spectrum of LP residual and (d) AP group delay.

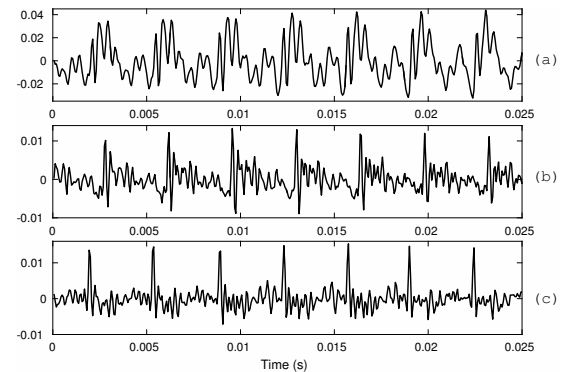


Figure 2: Illustration of residual signals: (a) Speech signal (b) LP residual and (c) AP residual.

set of AP coefficients (APCs) $\mathbf{w} = [w_1 w_2 \dots w_M]^T$, where M is the order of AP filter. The APCs \mathbf{w} has to be estimated for modeling the LP residual $y[n]$, which is an ill-posed problem as both the APCs \mathbf{w} and its input signal $x[n]$ are unknown. It requires some prior knowledge or assumption on either the filter or input signal, to solve the ill-posed APCs estimation problem. The estimation of APCs was done by assuming a dominant cumulant function of $x[n]$ [23], by enforcing a Laplacian distribution on $x[n]$ [24] or when $x[n]$ follows an arbitrary probability density function with known parameters [25]. But this prior knowledge are not available in case of natural signals, like speech.

In this work, we use the knowledge of speech production process for formulating constraints on $x[n]$. The voiced speech is produced by exciting the relatively unconstructed VTS with a quasi-periodic excitation signal, having significant energy only

at epochs and negligible energy elsewhere in a laryngeal cycle [26]. The excitation to VTS can be considered as a train of impulses, where energy is concentrated only at a few samples. Thus, we need to constrain the total energy of the input signal $x[n]$ to a few samples. Since the input and output signals of an AP filter hold the same energy, without loss of generality, the short-time segments of LP residual $y[n]$ of length N can be normalized to be unit energy signals. Thus the APCs estimation problem is formulated as : *Given the unit energy LP residual $y[n]$, estimate APCs \mathbf{w} , such that $x[n]$ has its unit energy concentrated to a few samples.* This can be achieved by minimizing the entropy of energy of $x[n]$. The sample-wise energy of $x[n]$ is expressed as: $e[n] = x^2[n]$ [27]. As $e[n] \geq 0, \forall n$ and $\sum_{n=1}^N e[n] = 1$, it can be viewed as a valid probability mass function . Hence the entropy of $e[n]$ is defined as [28]:

$$J(\mathbf{w}) = - \sum_{n=1}^N e[n] \log e[n] \quad (4)$$

And the APCs can be estimated as:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} J(\mathbf{w}) \quad (5)$$

The AP modeling strategy for phase spectrum by entropy minimization was proposed in [27]. In this work, we use the gradient descent algorithm with appropriately small step size, to minimize the entropy function $J(\mathbf{w})$ to obtain the APCs \mathbf{w} [27].

The group delay response of the estimated AP filter, for a short-time segment of speech is shown in Figure 1(d). It can be noticed that the peaks in group delay response coincide with the peaks in LP spectrum in Figure 1(b), demonstrating information about VTS resonances. The resultant AP residual is obtained by noncausal inverse filtering of LP residual $y[n]$ through estimated AP filter $H(z)$ [27], and is shown in Figure 2(c). The AP residual demonstrates unambiguous peaks at epochs, as opposed to multiple bipolar peaks around epochs in LP residual due to the presence of phase spectrum, as shown in Figure 2(b). The unmodeled phase spectrum of speech signals after LP analysis is modeled as the response of an AP filter, thereby generating the AP residual, which is a better representative of excitation source than the LP residual.

The unambiguous epochal information available in AP residual can be directly used as pitch markers, whereas an additional epoch extraction algorithm is required for pitch marking in prosody modification based on speech and LP residual. Also the envelope of magnitude spectrum of AP residual is nearly flat, similar to that of LP residual shown in Figure 1(b), as the AP filter does not modify the magnitude spectrum of signals. Thus prosody modification using AP residual will result in fewer spectral distortions than the TD-PSOLA. The use of AP residual nullifies the disadvantage of PSOLA algorithm (requirement of dedicated pitch marking) and secures the advantage associated with LP residual in prosody modification (nearly flat spectral envelope).

3. Prosody modification

The peaks in AP residual are identified using a criteria based on short-time energy. A sample of AP residual $x[n]$ at instant n is selected as a peak when it holds more than 80% of energy of its immediate neighborhood of 20 samples as [29]:

$$\frac{|x[n]|^2}{\sum_{k=-10, k \neq 0}^{10} |x[n+k]|^2} > 0.8 \quad (6)$$

The instants of selected peaks denote the epochs, and are used as pitch markers for prosody modification. As opposed to epoch extraction from LP residual, using dynamic programming algorithm with complex constraints, the AP residual-based epoch extraction is relatively simple.

Short-time overlapped windowing of AP residual is performed by placing windows centered at significant peaks in AP residual, which are identified as current pitch markers. Typically a window duration spans two pitch cycles. For pitch modification, the sequence of pitch markers is used to obtain a sequence of pitch mark intervals, constituting of the interval between two consecutive pitch markers. The sequence of pitch mark intervals is multiplied with the desired pitch modification factor to obtain a new sequence of pitch mark intervals, which is then used to obtain the instants of new pitch markers. The short-time segments of AP residual are realigned with respect to the new sequence of pitch marker instants and unique samples in each frame are retained to obtain the modified AP residual [17]. A segment of AP residual and its modified versions for 2 pitch modification factors are shown in Figure 3. The pitch modified speech signal is synthesized by filtering the modified AP residual through the cascade of AP and LP filters : $H(z)G(z)$, given in (3) and (2), respectively.

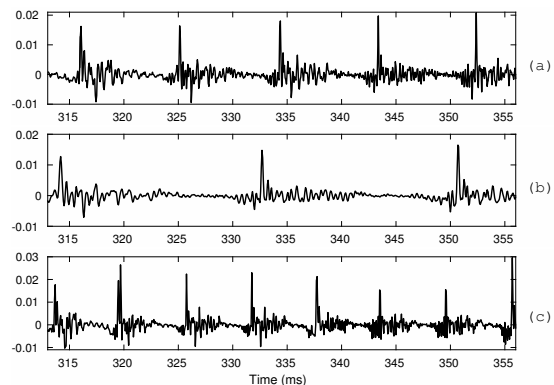


Figure 3: Illustration of pitch modification of AP residual: (a) Original AP residual (b) Pitch modified by a factor of 0.5 and (c) Pitch modified by a factor of 1.5.

For duration modification, the sequence of pitch mark intervals is resampled by the desired duration modification factor to obtain the new pitch mark interval sequence [17]. The instants of new pitch markers are obtained based on the new sequence of pitch mark intervals. Then the short-time segments of AP residual windowed around the old pitch markers are resampled with the desired duration modification factor. The resampling is done only on 80% of samples in a laryngeal cycle, while 20% samples around epochs are retained in the original form [17]. These resampled short-time segments are realigned with respect to the new pitch marker instants, and unique samples in each frame are retained to obtain the modified AP residual. The duration modified versions of AP residual corresponding to two modification factors are shown in Figure 4. By filtering the modified AP residual through the AP-LP cascade filter, duration modified speech signals are obtained. For synthesizing the duration modified speech signal, the filter coefficients are updated at the instants of window shift multiplied by the duration modification factor [17].

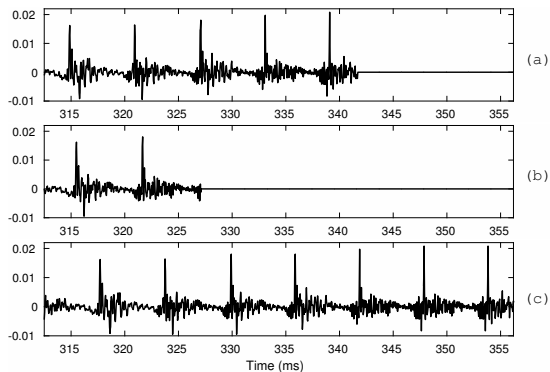


Figure 4: Illustration of duration modification of AP residual: (a) Original AP residual (b) Duration modified by a factor of 0.5 and (c) Duration modified by a factor of 1.5.

Table 1: MOS for duration modification strategies

Mod. factor	AP residual	LP residual	LP-PSOLA
0.4	3.87	3.71	3.67
0.5	4.16	3.88	3.82
0.8	4.52	4.56	4.42
1.25	4.57	4.59	4.48
1.5	4.81	4.78	4.13

4. Subjective Evaluation

Speech signals are sampled at 8 kHz and are segmented into short-time frames of 25 ms, shifted by 5 ms. LP analysis is performed on short-time frames of speech, and LPCs characterizing $G(z)$ and LP residual are obtained. AP modeling of short-time frames of LP residual is performed to obtain APCs characterizing $H(z)$ and the AP residual. The order of LP and AP analyses, M , is fixed at 14 [29]. The AP residual shows unambiguous peaks at epochs, which serve as pitch markers for voiced speech. The pitch markers for unvoiced speech are uniformly placed at every 5 ms interval. The duration and pitch of speech are modified by manipulating the sequence of pitch markers to obtain modified AP residuals, and the quality of prosody modification is evaluated using subjective experiments.

Speech utterances by a male and a female speaker from the ‘test’ subset of TIMIT database [30] are used for subjective evaluation. The duration and pitch of the utterances (3 utterances per speaker) are modified with 5 different modification factors as given in Table 1 and Table 2. Twenty-five normal hearing listeners, between the age of 20 and 30, participated in the subjective study. The speech files are played to the listeners in normal room environment using headphones. The listeners were asked to rate the perceptual quality of prosody modified speech utterances on a scale of 1 to 5, where 1 denotes unsatisfactory, 2 for poor, 3 for fair, 4 for good and 5 denotes excellent. The performance of duration and pitch modification strategies were evaluated based on mean opinion score (MOS) over all utterances of both male and female speakers. The performance of the proposed method using AP residual is compared with the LP residual-based method using epochs as pitch marks [17] and LP-PSOLA without knowledge of epochs [3]. The MOS for all three strategies for duration and pitch modifications are given in Table 1 and Table 2, respectively.

From Table 1 and Table 2, it can be seen that the proposed

Table 2: MOS for pitch modification strategies

Mod. factor	AP residual	LP residual	LP-PSOLA
0.5	3.77	3.81	3.89
0.8	4.52	4.58	4.36
1.25	4.42	4.46	4.23
1.5	4.39	4.36	3.79
2	3.88	3.84	3.26

strategy based on AP residual is delivering equivalent performance as that of LP residual-based method using epochs as pitch markers. Also, the proposed method provides better performance than the LP-PSOLA method operating without the knowledge of epochs. For small changes in pitch and duration (0.8 and 1.25), all the methods perform equivalently well. For duration compression of speech signals by a considerable factor, the performance of AP based method is better than all other strategies, due to the marginal information loss happening during down-sampling of AP residual. The AP residual has its prominent energy centered around epochs (which are not down-sampled) and negligible energy elsewhere in a laryngeal cycle (which are down-sampled), resulting in little information loss. In case of pitch reduction by a considerable factor, the AP based method becomes slightly inferior to LP residual-based method utilizing epochs, because of the greater duration between successive peaks in modified AP residual. This causes minor discontinuity in synthesis resulting in perceivable distortions.

The proposed method for prosody modification had successfully utilized the epochal information available in AP residual signal, which is obtained by modeling the magnitude and phase spectra of speech signals. Also, the AP residual-based prosody modification induces fewer spectral distortions than TD-PSOLA, due to the nearly flat magnitude spectrum. The samples of AP residual are maximally independent, and hence are robust to time domain manipulations like resampling. Also, the proposed method did not require an accurate pitch marking algorithm. However, the additional computation required for AP modeling could be a potential disadvantage for prosody modification in real-time applications.

5. Conclusions

In this paper, the significance of modeling the phase spectrum of speech signals in obtaining a true representative of excitation source for prosody modification was presented. The phase spectrum was modeled as the response of an allpass (AP) filter, whose output was chosen to be the LP residual. The estimation of AP filter was done by minimizing an entropy based objective function using gradient descent algorithm. The resultant AP residual held maximally independent samples, a nearly flat magnitude spectrum and exhibited unambiguous peaks at epochs. The unambiguous information about epochs in AP residual were used as pitch markers for prosody modification. Hence the AP based method did not require an accurate pitch marking algorithm, which is not the case with other PSOLA techniques. Also, the nearly flat spectral envelope of AP residual resulted in fewer spectral distortions in prosody modified speech signals, in comparison with TD-PSOLA algorithm. The subjective evaluation of prosody modified speech signals synthesized from AP residual revealed the efficacy of the proposed technique in comparison with other state-of-the-art methods.

6. References

- [1] T. F. Quatieri and R. J. McAulay, "Shape invariant time-scale and pitch modification of speech," *IEEE Transactions on Signal Processing*, vol. 40, no. 3, pp. 497–510, Mar 1992.
- [2] D. H. Klatt, "Review of text-to-speech conversion for English," *The Journal of the Acoustical Society of America*, vol. 82, pp. 737–793, 1987.
- [3] E. Moulines and F. Charpentier, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, no. 5, pp. 453 – 467, 1990.
- [4] M. Portnoff, "Time-scale modification of speech based on short-time Fourier analysis," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 3, pp. 374–390, Jun 1981.
- [5] D. Childers, K. Wu, D. Hicks, and B. Yegnanarayana, "Voice conversion," *Speech Communication*, vol. 8, no. 2, pp. 147 – 158, 1989.
- [6] T. Dutoit and H. Leich, "MBR-PSOLA: Text-to-speech synthesis based on an MBE re-synthesis of the segments database," *Speech Communication*, vol. 13, no. 3, pp. 435 – 440, 1993.
- [7] W. Verhelst and M. Roelands, "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'93)*, vol. 2, Apr 1993, pp. 554–557 vol.2.
- [8] R. Veldhuis, "Consistent pitch marking," in *International Conference on Spoken Language Processing*, Oct 2000, pp. 207–210.
- [9] Y. Laprie and V. Colotte, "Automatic pitch marking for speech transformations via TD-PSOLA," in *9th European Signal Processing Conference*, Sep 1998, pp. 1–4.
- [10] W. Mattheyse, W. Verhelst, and P. Verhoeve, "Robust pitch marking for prosodic modification of speech using TD-PSOLA," 2006, pp. 43–46.
- [11] C.-Y. Lin and J.-S. R. Jang, "A two-phase pitch marking method for TD-PSOLA synthesis," in *INTERSPEECH - ICSLP*, Oct 2004, pp. 1189–1192.
- [12] V. Colotte and Y. Laprie, "Higher precision pitch marking for TD-PSOLA," in *11th European Signal Processing Conference*, Sep 2002, pp. 1–4.
- [13] T. Ewender and B. Pfister, "Accurate pitch marking for prosodic modification of speech segments," in *INTERSPEECH*, Sep 2010, pp. 178–181.
- [14] A. Chalamandaris, P. Tsiakoulis, S. Karabetsos, and S. Raptis, "An efficient and robust pitch marking algorithm on the speech waveform for TD-PSOLA," in *IEEE International Conference on Signal and Image Processing Applications (ICSIPA'09)*, Nov 2009, pp. 397–401.
- [15] J. P. Cabral and L. C. Oliveira, "Pitch-synchronous time-scaling for prosodic and voice quality transformations," in *INTERSPEECH*, 2005, pp. 1137–1140.
- [16] F. M. G. de los Galanes, M. H. Savoji, and J. M. Pardo, "New algorithm for spectral smoothing and envelope modification for LP-PSOLA synthesis," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'94)*, vol. i, 1994, pp. I/573–I/576 vol.1.
- [17] K. S. Rao and B. Yegnanarayana, "Prosody modification using instants of significant excitation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 3, pp. 972–980, May 2006.
- [18] Y. Stylianou, "Applying the harmonic plus noise model in concatenative speech synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 1, pp. 21–29, Jan 2001.
- [19] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 323–332, May 1999.
- [20] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab, *Signals and systems*, 2nd ed. Upper Saddle River, NJ, USA: Pearson Education Inc., 1997.
- [21] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, Apr 1975.
- [22] K. S. R. Murty, V. Boominathan, and K. Vijayan, "Allpass modeling of LP residual for speaker recognition," in *International Conference on Signal Processing and Communications*, Jul 2012, pp. 1–5.
- [23] C.-Y. Chi and J.-Y. Kung, "A new identification algorithm for allpass systems by higher-order statistics," in *Signal Processing*, vol. 41, January 1995, pp. 239–256.
- [24] F. J. Breidt, R. A. Davis, and A. A. Trindade, "Least absolute deviation estimation for all-pass time series models," in *Annals of statistics*, vol. 29, 2001, pp. 919–946.
- [25] B. Andrews, R. A. Davis, and F. J. Breidt, "Maximum likelihood estimation for all-pass time series models," in *Journal of Multivariate Analysis*, vol. 97, August 2006, pp. 1638–1659.
- [26] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1978.
- [27] K. Vijayan and K. S. R. Murty, "Analysis of phase spectrum of speech signals using allpass modeling," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2371–2383, Dec 2015.
- [28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, ser. Telecommunications and signal processing. Wiley-Interscience, 2006.
- [29] K. Vijayan and K. S. R. Murty, "Epoch extraction by phase modelling of speech signals," *Circuits, Systems, and Signal Processing*, pp. 1–26, 2015.
- [30] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus," 1993.