Allpass Modeling of Phase Spectrum of Speech Signals for Formant Tracking

Karthika Vijayan*, K. Sri Rama Murty† and Haizhou Li*
* Department of Electrical and Computer Engineering
National University of Singapore, Singapore
E-mails: {vijayan.karthika, haizhou.li}@nus.edu.sg
† Department of Electrical Engineering
Indian Institute of Technology Hyderabad, India
E-mail: ksrm@iith.ac.in

***Abstract*—Formant tracking is a very important task in speech applications. Most of the current formant tracking methods bank on peak picking from linear prediction (LP) spectrum of speech, which suffers from merged/spurious peaks in LP spectra, resulting in unreliable formant candidates. In this paper, we present the significance of phase spectrum of speech in refining the formant candidates from LP analysis. The short-time phase spectrum of speech is modeled as phase response of an allpass (AP) system, where the coefficients of AP system are initialized with LP coefficients and estimated with an iterative procedure. This technique refines the initial formants from LP analysis using phase spectrum of speech through an AP analysis, thereby accomplishing fusion of information from magnitude and phase spectra. The group delay of the resultant AP system exhibits unambiguous peaks at formants and, delivers reliable formant candidates. The formant trajectories obtained by selection of formants from these candidates are reported to be more accurate than those obtained from LP analysis. The fused information from magnitude and phase spectra has rendered relative improvements of 25%, 15% and 18% in tracking accuracy of first, second and third formants, respectively, over those from magnitude spectrum alone.**

Index Terms: Formant tracking, Phase spectrum, Allpass modeling, Fusion of information.

I. Introduction

Formants are defined as peaks in the spectrum of speech sounds [1]. As they are the high energy content in spectrum, formants are relatively robust to various distortions, and are used for speech enhancement [2]. Acoustically, formants represent resonances of the human vocal tract system (VTS) and there exist direct correspondence between formants and VTS parameters. This correspondence has been studied for speech recognition [3], speaker recognition [4], speech analysis [3], speech synthesis [1], [5]–[7], voice conversion [8], [9], etc. However, difficulties in obtaining accurate formant trajectories had limited their use to study of speech sounds and, restricted the utilization of formant-based features in mainstream speech processing.

As formants are defined as peaks in speech spectrum, the most straightforward way to identify formants is peak picking from short-time magnitude spectra of speech signals. Peak picking from cepstrally smoothed magnitude spectrum and linear prediction (LP) spectrum were attempted to identify formants [10]–[12]. However, all formants may not be manifested as observable peaks in smoothed speech spectrum. Hence, solving for roots of LP polynomial to compute poles of the system was employed for identifying formants, that are not readily observable from the LP spectrum [13]. Later, carefully designed continuity constraints were employed in a dynamic programming framework to obtain smooth formant trajectories [3], [14], [15]. Instead of using the LP spectrum or LP polynomial directly, several statistical and adaptive methodologies including hidden Markov models [16]–[18], mixture models [19], [20], particle filters [21]–[23] and Kalman filters [24], [25] were employed for formant tracking. Notice that, most of these methods rely on magnitude spectral information in speech signals.

In addition to magnitude spectrum, phase spectrum was employed in formant tracking by the use of group delay functions. The computation of group delay of speech is affected by several sharp and spurious peaks. Smoothing of group delay function was attempted by conditioning it with magnitude spectrum, or by constraining the analysis to within unit circle in z-plane [26], [27]. The LP analysis was also employed for computation of group delay from LP residual to represent phase spectrum [28], [29]. These methods essentially impose a minimum phase assumption on speech signals for computation of group delay, which is not necessarily a valid assumption.

Another class of techniques employed for formant tracking involves segmentation of frequency range of speech signals, instead of short-time analysis. Parallel resonator models, filterbank analysis, time-varying adaptive filters, etc. were employed for the frequency segmentation of speech for formant tracking [30]–[34].

In this paper, we attempt to perform formant tracking in frequency domain using short-time analysis of speech, without imposing any minimum phase assumption on VTS. We consider the VTS to be a mixed phase system and utilize the short-time magnitude and phase spectra of speech, represented by LP (minimum phase) and AP (allpass) analyses, for formant tracking. The phase spectrum of speech is modeled as the phase response of an AP system, where the coefficients of the AP system (APCs) are iteratively estimated. In this iterative process, the APCs are initialized with the coefficients of LP

system obtained from LP analysis of short-time segment of speech. Thus, the magnitude spectral information embedded in LP system is used as a starting point to model phase spectrum into an AP system. The group delay of the resultant AP system exhibits unambiguous peaks at formants. The smoothed formant trajectories are obtained by choosing formants from candidates formed from poles of the AP system function, using dynamic programming based on continuity constraints. It is observed that the proposed use of phase spectrum has delivered enhanced formant tracking performance when compared to methods relying only on magnitude spectrum from LP analysis.

The rest of the paper is organized as follows: Section II discusses the AP modeling of phase spectrum of speech and its usefulness in identifying formants. The formant tracking methodology using phase spectrum of speech is explained in Section III. Formant tracking performances of different methods are evaluated in Section IV. In Section V, we summarize the contributions of this paper.

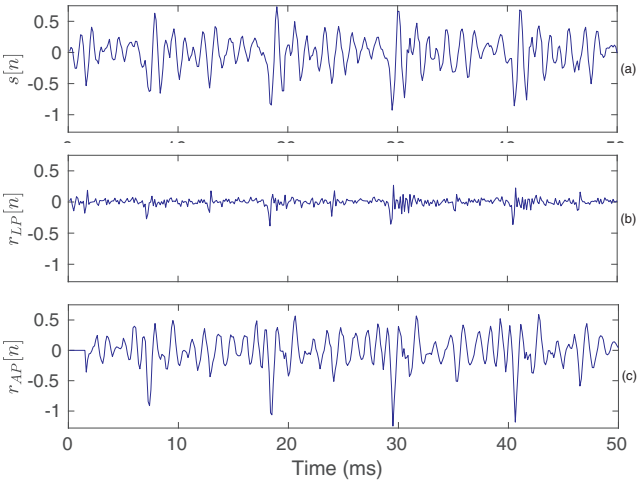

Fig. 1. Signals involved in LP and AP analyses: (a) A segment of voiced speech, (b) LP residual from 12^{th} order LP analysis and (c) AP residual from 14^{th} order AP analysis.

II. Modeling phase spectrum for formant tracking

Speech signal is the output of a time-varying VTS excited with a time-varying glottal excitation. As formants are manifestation of resonances of the VTS, efficient source-system separation from speech signals can provide ways to compute formants. For source-system separation, speech can be analyzed using short-time Fourier transform. The Fourier transform of a short-time segment of discrete-time speech signal, $s[n]$ is given by,

$$S(j\omega) = \sum_{n=0}^{N-1} s[n]e^{-j\omega n} = |S(j\omega)|e^{j\angle S(j\omega)} \quad (1)$$

where, $|S(j\omega)|$ and $\angle S(j\omega)$ are the magnitude and phase spectrum, respectively. A segment of voiced speech is shown in Fig. 1(a) and corresponding magnitude and phase spectra are given in Fig. 2(a) and Fig. 2(b), respectively. In general, the phase and magnitude spectra are required to completely characterize $S(j\omega)$.

LP analysis is a prominent method for characterizing the magnitude spectral envelope, as it models the VTS as an autoregressive and all-pole model [11]. The error incurred in LP analysis, termed as LP residual, represents the excitation signal. The LP analysis essentially performs source-system separation from speech signals. The magnitude responses of LP systems for two model orders are shown in Fig. 2(c) and Fig. 2(e) and, the LP residual is shown in Fig. 1(b). The peaks in the LP magnitude response can be deduced as short-time formants, and significant peaks in LP residual indicate the glottal closure instants [35].

Though magnitude spectrum is widely used in speech applications, the phase spectrum is usually ignored. In this work, we present the effectiveness of phase spectrum in identification of formants. A procedure to model the phase spectrum of speech signals was proposed in [36]. The magnitude spectrum of speech is deemphasized to highlight the phase spectral characteristics, and are modeled into an AP system response. The steps involved in AP modeling of phase spectrum are [36]:
1) A signal is generated from speech signal by removing the magnitude spectrum and preserving the phase spectrum as,

$$y[n] = \mathcal{F}^{-1}\left\{\frac{S(j\omega)}{|S(j\omega)|}\right\} \quad (2)$$

where $\mathcal{F}^{-1}$ is the inverse Fourier transform. This signal, termed as phase signal $y[n]$ has uncorrelated, but not independent samples.
2) The phase signal is considered as the output of an allpass system, excited with independent and identically distributed (i.i.d) non-Gaussian input sequence, $x[n]$.
3) The sequence $x[n]$ is considered as a representative of glottal excitation. For voiced speech, significant excitation happens only at glottal closure instants and hence, the $x[n]$ has its energy, $e[n] = x^2[n]$, concentrated to a few number of samples.
4) The AP system has its poles and zeros located at conjugate reciprocal locations. Hence, the numerator and denominator of the AP system function are characterized by the same set of coefficients $\mathbf{w} = [w_1 w_2 ... w_M]^T$.
5) Estimate the APCs of an M^{th} order AP system, $\mathbf{w}$, by minimizing the entropy of energy of $x[n]$, $J(\mathbf{w}) = -\sum_{n=1}^{N} e[n]\ log\ e[n]$. The APCs can be estimated as:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} J(\mathbf{w}) \quad (3)$$

The algorithm for estimation of APCs $\mathbf{w}$, by minimizing $J(\mathbf{w})$ using gradient descent optimization is given in Algorithm. 1.

A. Significance of phase spectrum

The LP analysis models the VTS as a minimum phase all-pole system and characterizes the magnitude spectral envelope of speech sigals. Hence, it fails to represent the zeros of the system function of VTS. Also, the unmodeled phase

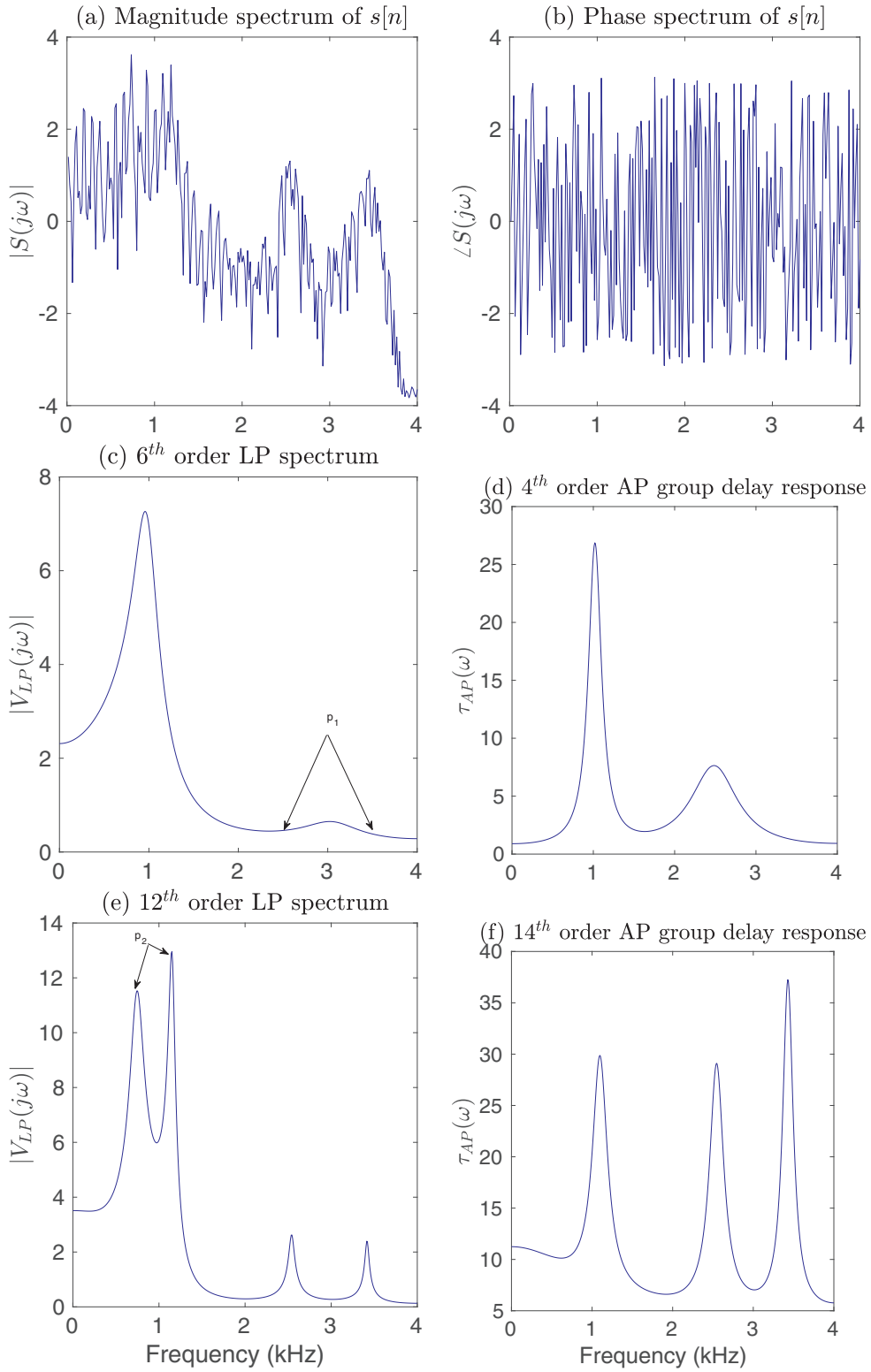

Fig. 2. Illustration of frequency responses of LP and AP analyses.

Algorithm 1 AP modeling using gradient descent algorithm [36].

1: Initialize $\mathbf{w_0} \leftarrow rand(M,1)$ and $l \leftarrow 1$
2: **while** $J(\hat{\mathbf{w}}_{l-1}) - J(\hat{\mathbf{w}}_l) > \epsilon$, where ϵ is chosen as 10^{-6} **do**
3: Compute gradient as

$$\nabla J(\mathbf{w}) = \frac{\partial J(\mathbf{w})}{\partial e[n]} \frac{\partial e[n]}{\partial x[n]} \nabla x[n]$$
$$= -\sum_{n=1}^{N} (1 + \log e[n]) (2x[n]) \nabla x[n]$$

where

$$\nabla x[n] = -\sum_{k=1}^{M} w_k \nabla x[n+k] - \mathbf{x}[n+1] + \mathbf{y}[n+M-1]$$

and

$$\mathbf{x}[n+1] = [x[n+1]\ x[n+2]\ ....\ x[n+M]]^T,$$
$$\mathbf{y}[n+M-1] = [y[n+M-1]\ y[n+M-2]\ ....\ y[n]]^T$$
$$\nabla = [\tfrac{\partial}{\partial w_1} \tfrac{\partial}{\partial w_2} \cdots \tfrac{\partial}{\partial w_M}]^T.$$

4: Update as $\hat{\mathbf{w}}_l = \hat{\mathbf{w}}_{l-1} - \eta \nabla J(\mathbf{w}) \big|_{\mathbf{w}=\hat{\mathbf{w}}_{l-1}}$, where step-size is chosen as $\eta = 0.005$
5: $l \leftarrow l + 1$
6: **end while**

spectrum gets reflected in the LP residual, forming multiple spurious peaks around glottal closure instants and preventing the residual signal from being a true representative of the excitation (See Fig. 1(b)). In addition, the LP magnitude response tends to exhibit single peak corresponding to two closely spaced peaks in the original spectrum (marked as p_1 in Fig. 2(c)), and multiple peaks corresponding to single high bandwidth spectral peaks (marked as p_2 in Fig. 2(e)). These scenarios are referred to as merged and spurious peaks scenarios. Thus unique identification of formants from LP analysis is not straightforward as the source-system separation is not fully accomplished.

A possible solution is to capture the unmodeled phase spectrum from LP analysis to estimate a better representation of VTS and excitation signal. The mixed phase VTS can be effectively represented in terms of its minimum phase and allpass components using LP and AP modelings. The group delay response of AP system and the AP residual obtained by modeling the phase spectrum of speech signal in Fig. 1(a) are shown in Fig. 2(f) and Fig. 1(c), respectively. The AP residual exhibits single peak at each glottal closure instant, thereby representing the excitation signal unambiguously. When the model order is chosen appropriately, the group delay response of AP system exhibits explicit distinguishable peaks at formants. Notice that, the group delay of LP system also exhibits sharper peaks than LP magnitude response. However, this group delay is affected by spurious/merged peaks scenarios as they are dictated by the poles of the LP system.

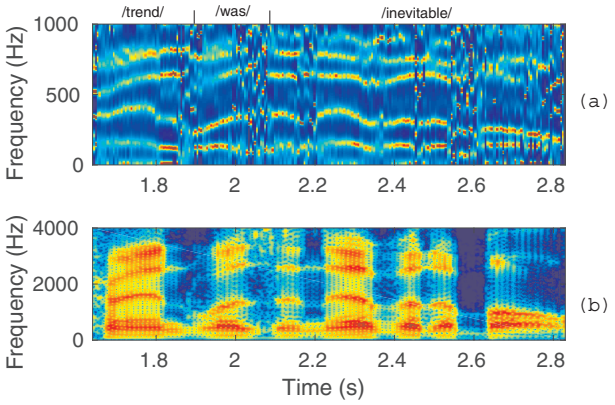

Fig. 3. Groupdelaygram and Spectrogram of speech signal.

To further illustrate the effectiveness of AP modeling, we plot the groupdelaygram from AP systems estimated from overlapping short-time segments of speech signal (See Fig. 3). The formant tracks are well preserved in the groupdelaygram as in the speech spectrogram, indicating the potential use of AP systems modeling phase spectrum of speech in formant tracking.

The AP analysis, together with LP analysis, can be utilized to deliver a near-complete source-system separation from speech signals. Formant tracking using the AP system makes use of magnitude and phase spectra of speech, instead of magnitude spectrum alone in LP analysis.

III. Formant Tracking Using Phase Spectrum

In order to use the information in magnitude and phase spectra for formant tracking, we perform LP and AP analyses in succession and use the poles of AP system to identify short-time formant candidates. The LP analysis is performed to obtain coefficients of the LP system (LPCs). Later AP analysis is conducted, in which we propose to initialize the APCs in Step 1 of Algorithm 1 with the LPCs. Thus the initial information about formants embedded in LP system is made use of in AP estimation, which got refined to new formant information in the AP system. Root solving of denominator polynomial of the AP system provides frequencies and bandwidths associated with the poles, which form new short-time formant candidates that are more reliable than those obtained from LP system alone.

This strategy, which we call LP+AP analysis, delivers $M/2$ formant candidates at the maximum for model order M. Three formant values have to be chosen from the candidates, forming trajectories for first three formants, namely, F_1, F_2 and F_3. This is accomplished using a dynamic programming (DP) algorithm developed in [15], in-line with [14]. The DP utilizes complex poles of AP system (real roots are neglected) to compute the frequency f_i and bandwidth b_i of the formant candidates. As the first 3 formants are to be tracked, certain parameters associated with them have to be initialized as: (1) boundary conditions: $100 < F_1 < 1500, 500 < F_2 < 3500, 1000 < F_3 < 4500$ and (2) the neutral values of VTS

TABLE I
FORMANT TRACKING PERFORMANCE IN TERMS OF MEAN ABSOLUTE DEVIATION (IN %) OF ESTIMATED FORMANTS FROM THEIR REFERENCE VALUES.

Classes	LP			AP			LP+AP		
	F_1	F_2	F_3	F_1	F_2	F_3	F_1	F_2	F_3
Vowels	17.26	10.92	8.97	24.78	14.27	17.13	**14.49**	**9.36**	**8.68**
Semi-vowels	44.41	44.48	36.29	34.11	29.46	27.49	**31.74**	**27.12**	**26.16**
Nasals	44.55	32.76	17.91	25.92	20.37	10.35	**20.71**	**18.90**	**10.37**
Average	35.41	29.39	20.05	28.27	21.37	18.32	**22.31**	**18.46**	**15.07**

for first 3 formants: $F_{n1} = 500, F_{n2} = 1500, F_{n3} = 2500$ [15], [24]. The cost function for the DP is formulated for each node of formants, d and each frame of speech indexed by p as [15]:

$$C(p,d) = C_l(p,d) + \min_m \{C_t((p,d),(p-1,m)) + C(p-1,m)\} \quad (4)$$

where nodes of formants are formed by the $(M/2)C_3$ combinations of formant candidates at each frame index p. The local and translational costs involved in Equation (4) are defined as [15]:

$$C_l(p,d) = \sum_j \alpha_j b_j^2 + \beta_j \frac{|f_j - F_{nj}|}{F_{nj}}, \quad (5a)$$

$$C_t((p,d),(p-1,m)) = \sum_j \gamma_j (f_j(p) - f_j(p-1))^2. \quad (5b)$$

$\forall j = 1,2,3$ and m represents the node with the lowest cost in the preceding frame indexed as $p-1$. The constants α, β and γ, control the relative weighting of different cost functions for different formants, whose values are carefully decided.

The DP searches for the lowest cost path, minimizing the cost function $C(p,d)$, and provides the estimates of F_1, F_2 and F_3 for each frame of speech. The search is carried out on vowel-like regions and nasal regions separately [15]. In order to avoid ambiguities or discontinuities caused by outliers or missing formant points, a moving average smoothing is performed. Thus smooth formant tracks for F_1, F_2 and F_3 are obtained. The speech spectrogram and estimated formant tracks from various analyses are shown in Fig. 4. The LP+AP analysis (APCs initialized with LPCs) provided reliable formant tracks in comparison to those obtained from individual LP and AP analysis (APCs with random initialization).

IV. EXPERIMENTAL EVALUATION

To evaluate the formant tracking performance of various strategies, we use the VTR-Formants database [37]. This database consists of speech signals, together with manually marked formant values. The dataset chosen for evaluation consists of 500 utterances from 8 female and 16 male speakers. The utterances have an average duration of 4 seconds, and are sampled at 8 kHz. The utterances are segmented into frames of 25 ms duration, with a time shift of 10 ms. The frames are selected from voiced speech using an energy-based voicing decision. The LP, AP and LP+AP analyses are performed on these short-time frames. The order M for all the analyses is fixed as 14 [38].

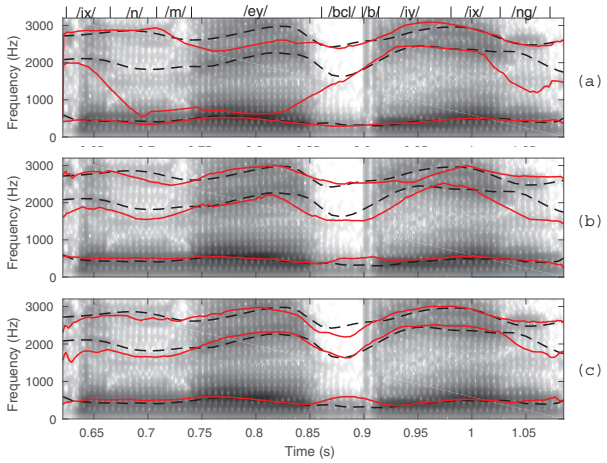

Fig. 4. Formant tracks obtained (a) LP, (b) AP and (c) LP+AP analyses (depicted by solid lines). Dotted lines represent reference values of formants from VTR-Formants database.

In the evaluation of formant tracking performance, we have excluded obstruent speech sounds, as formants are mostly meaningful in sonorant speech [15]. The evaluation of formant tracking is carried out on 3 classes of speech sounds separately, as, vowels, semivowels and nasals. The evaluation results, in terms of mean of absolute values of deviations of the estimated formants from the respective manually marked reference values in VTR database, are given in Table. I. The fusion of information from magnitude and phase spectra of speech has proven advantageous in formant tracking. The proposed LP+AP modeling strategy outperforms formant tracking by either LP or AP modeling alone, in all speech classes.

The AP modeling outperforms LP analysis in formant tracking from nasals and diphthongs. The formant tracks from LP analysis largely deviates from their reference values for sounds /n/, /m/ and /ng/, as illustrated in Figure 4(a). For the sound /ey/ shown in Figure 4(a), the LP analysis places 2 peaks corresponding to the first formant with wider bandwidth, causing a large error in the estimated second formant. On the other hand, LP analysis outperforms AP modeling in formant tracking from vowels. This can be observed from Figure 4(a) and Figure 4(b) for sounds /iy/ and /ix/. Also the AP modeling performance suffers in closure phase of stop sounds. The LP+AP strategy nullifies the drawbacks of both AP and LP analyses and consistently delivers best performance for all classes of speech sounds. The LP+AP strategy had provided relative improvements of 25%, 15% and 18% in

average formant tracking performance for the first 3 formants respectively, with respect to the LP-based method.

V. Conclusions

We proposed a strategy to improve the formant tracking performance by LP analysis, using phase spectrum of speech signals. The short-time phase spectrum of speech was modeled as the response of an AP system, and the initial values of AP coefficients in an iterative AP modeling algorithm were substituted with the LP coefficients. Thus the fusion of information from magnitude spectrum (captured in LPCs) and phase spectrum (captured in APCs) was accomplished. The formant tracking was performed using a DP, which selected the short-time estimates of formants from the candidates obtained by root solving of AP polynomial. The DP also imposed continuity constraints on short-time formant estimates to acquire smooth formant tracks. The formant tracks thus obtained from the proposed LP+AP modeling strategy outperformed those obtained from either LP or AP modeling alone, demonstrating the superiority of fusion of information from magnitude and phase spectra.

Acknowledgment

The first and third authors would like to acknowledge the research support by Ministry of Education, Singapore, AcRF Tier 1 NUS Start-up Grant FY2016 for the project 'Non-parametric approach to voice morphing'.

References

[1] G. Fant, *Acoustic theory of speech production*. Mouton, The Hague, 1960.

[2] Q. Yan, S. Vaseghi, E. Zavarehei, B. Milner, J. Darch, P. White, and I. Andrianakis, "Formant tracking linear prediction model using HMMs and Kalman filters for noisy speech processing," *Computer Speech & Language*, vol. 21, no. 3, pp. 543–561, Jul. 2007.

[3] L. Welling and H. Ney, "Formant estimation for speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 1, pp. 36–48, Jan 1998.

[4] J. J. Wolf, "Efficient acoustic parameters for speaker recognition," *The Journal of the Acoustical Society of America*, vol. 51, no. 6B, pp. 2044–2056, 1972.

[5] J. L. Flanagan, J. B. Allen, and M. A. Hasegawa-Johnson, *Speech Analysis Synthesis and Perception*, 3rd ed. Springer-Verlag, 2008.

[6] D. Klatt, "Software for a cascade/parallel formant synthesizer," *The Journal of the Acoustical Society of America*, vol. 67, no. 3, pp. 971–995, 1980.

[7] N. Pinto, D. G. Childers, and A. Lalwani, "Formant speech synthesis: improving production quality," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 12, pp. 1870–1887, Dec 1989.

[8] D. Rentzos, S. Vaseghi, Q. Yan, C. H. Ho, and E. Turajlic, "Probability models of formant parameters for voice conversion," in *EUROSPEECH*, 2003, pp. 2405–2408.

[9] H. Mizuno and M. Abe, "Voice conversion algorithm based on piecewise linear conversion rules of formant frequency and spectrum tilt," *Spch. Comm.*, vol. 16, no. 2, pp. 153–164, 1995.

[10] R. W. Schafer and L. R. Rabiner, "System for automatic formant analysis of voiced speech," *The Journal of the Acoustical Society of America*, vol. 47, no. 2B, pp. 634–648, 1970.

[11] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *The Journal of the Acoustical Society of America*, vol. 50, no. 2B, pp. 637–655, 1971.

[12] J. Markel, "Application of a digital inverse filter for automatic formant and F0 analysis," *IEEE Transactions on Audio and Electroacoustics*, vol. 21, no. 3, pp. 154–160, Jun 1973.

[13] S. McCandless, "An algorithm for automatic formant extraction using linear prediction spectra," *IEEE Trans. on Acoustics, Spch. and Signal Proc.*, vol. 22, no. 2, pp. 135–141, Apr 1974.

[14] D. Talkin, "Speech formant trajectory estimation using dynamic programming with modulated transition costs," *The Journal of the Acoustical Society of America*, vol. 82, no. S1, pp. S55–S55, 1987.

[15] K. Xia and C. Y. Espy-Wilson, "A new strategy of formant tracking based on dynamic programming." in *INTERSPEECH*, 2000, pp. 55–58.

[16] G. Kopec, "Formant tracking using hidden markov models and vector quantization," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 4, pp. 709–729, Aug 1986.

[17] D. Toledano, J. Villardebo, and L. Gomez, "Initialization, training, and context-dependency in HMM-based formant tracking," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 511–523, March 2006.

[18] M. Lee, J. van Santen, B. Mobius, and J. Olive, "Formant tracking using context-dependent phonemic information," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 741–750, Sept 2005.

[19] P. Zolfaghari, S. Watanabe, A. Nakamura, and S. Katagiri, "Bayesian modelling of the speech spectrum using mixture of gaussians," in *IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP'04)*, vol. 1, May 2004, pp. I–553–6 vol.1.

[20] E. Ozkan, I. Ozbek, and M. Demirekler, "Dynamic speech spectrum representation and tracking variable number of vocal tract resonance frequencies with time-varying dirichlet process mixture models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 8, pp. 1518–1532, Nov 2009.

[21] Y. Zheng and M. Hasegawa-Johnson, "Formant tracking by mixture state particle filter," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'04)*, vol. 1, May 2004, pp. I–565–8 vol.1.

[22] K. Kalgaonkar and M. Clements, "Vocal tract area based formant tracking using particle filter," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'08)*, Mar 2008, pp. 3405–3408.

[23] Y. Shi and E. Chang, "Spectrogram-based formant tracking via particle filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, vol. 1, 2003, pp. I–168–I–171 vol.1.

[24] L. Deng, L. Lee, H. Attias, and A. Acero, "Adaptive Kalman filtering and smoothing for tracking vocal tract resonances using a continuous-valued hidden dynamic model," *IEEE Trans. on Audio, Spch., and Lang. Proc.*, vol. 15, no. 1, pp. 13–23, 2007.

[25] D. D. Mehta, D. Rudoy, and P. J. Wolfe, "Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking," *The Journal of the Acoustical Society of America*, vol. 135, no. 5, pp. 3128–3128, 2014.

[26] H. A. Murthy and B. Yegnanarayana, "Group delay functions and its applications in speech technology," *Sadhana*, vol. 36, no. 5, pp. 745–782, 2011.

[27] B. Bozkurt, L. Couvreur, and T. Dutoit, "Chirp group delay analysis of speech signals," *Speech Communication*, vol. 49, no. 3, pp. 159–176, 2007.

[28] B. Yegnanarayana, "Formant extraction from linear prediction phase spectra," *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1638–1640, 1978.

[29] M. K. Dhananjaya Gowda, Jouni Pohjalainen and P. Alku, "Robust formant detection using group delay function and stabilized weighted linear prediction," in *INTERSPEECH*, 2004, pp. 49–53.

[30] A. Potamianos and P. Maragos, "Speech formant frequency and bandwidth tracking using multiband energy demodulation," *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3795–3806, 1996.

[31] K. Mustafa and I. C. Bruce, "Robust formant tracking for continuous speech with speaker variability," *IEEE Trans. on Audio, Spch., and Lang. Proc.*, vol. 14, no. 2, pp. 435–444, Mar 2006.

[32] A. Rao and R. Kumaresan, "On decomposing speech into modulated components," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 240–254, May 2000.

[33] J. Vargas and S. Mclaughlin, "Cascade prediction filters with adaptive zeros to track the time-varying resonances of the vocal tract," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 1–7, Jan 2008.

[34] L. Deng and I. Kheirallah, "Dynamic formant tracking of noisy speech using temporal analysis on outputs from a nonlinear cochlear model,"

IEEE Transactions on Biomedical Engineering, vol. 40, no. 5, pp. 456–467, May 1993.

[35] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1978.

[36] 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.

[37] L. Deng, X. Cui, R. Pruvenok, Y. Chen, and S. Momen, "A database of vocal tract resonance trajectories for research in speech processing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'06)*, vol. 1, May 2006, pp. I–I.

[38] K. Vijayan and K. S. R. Murty, "Epoch extraction by phase modelling of speech signals," *Circuits, Systems, and Signal Processing*, pp. 1–26, 2015.