Research Article

A Ramp Cosine Cepstrum Model for the Parameter Estimation of Autoregressive Systems at Low SNR

Shaikh Anowarul Fattah,[1, 2] Wei-Ping Zhu,[2] and M. Omair Ahmad[2]

[1] *Department of Electrical Engineering, Princeton University, Engineering Quadrangle, Olden Street, Princeton, NJ 08544, USA*
[2] *Department of Electrical and Computer Engineering, Concordia University, 1455 De Maisonneuve Blvd. W., Montreal, QC, Canada H3G 1M8*

Correspondence should be addressed to Shaikh Anowarul Fattah, sfattah@princeton.edu

A new cosine cepstrum model-based scheme is presented for the parameter estimation of a minimum-phase autoregressive (AR) system under low levels of signal-to-noise ratio (SNR). A ramp cosine cepstrum (RCC) model for the one-sided autocorrelation function (OSACF) of an AR signal is first proposed by considering both white noise and periodic impulse-train excitations. Using the RCC model, a residue-based least-squares optimization technique that guarantees the stability of the system is then presented in order to estimate the AR parameters from noisy output observations. For the purpose of implementation, the discrete cosine transform, which can efficiently handle the phase unwrapping problem and offer computational advantages as compared to the discrete Fourier transform, is employed. From extensive experimentations on AR systems of different orders, it is shown that the proposed method is capable of estimating parameters accurately and consistently in comparison to some of the existing methods for the SNR levels as low as -5 dB. As a practical application of the proposed technique, simulation results are also provided for the identification of a human vocal tract system using noise-corrupted natural speech signals demonstrating a superior estimation performance in terms of the power spectral density of the synthesized speech signals.

1. Introduction

The parameter estimation of autoregressive (AR) systems under noisy conditions has been extensively studied in areas of signal processing, communication, and control. For example, estimating the AR or linear predictive coding (LPC) parameters of a vocal tract (VT) system from an observed noisy speech plays an important role in speech coding, synthesis, and recognition [1]. Numerous system identification methods have been developed for both noise-free and noisy AR systems. The maximum likelihood (ML) methods are asymptotically consistent but their convergence performance relies heavily on the initialization process of the methods [2, 3]. In [3], Xie and Leung have proposed a genetic algorithm to be employed to solve the ML estimation problem at a low SNR, where they consider an AR system driven by chaos. The Yule-Walker (YW) methods have been widely employed to identify the AR systems [2].

The estimation performance of noise compensation-based identification schemes, such as the low-order Yule-Walker (LOYW) method, depend heavily on the accuracy of *a priori* knowledge of the noise corrupting the signal [2]. Although the high-order Yule-Walker (HOYW) method does not require *a priori* estimate of the noise variance, it suffers from a singularity problem and has a large estimation variance [4]. To reduce the estimation variance, a least-squares HOYW (LSYW) method can be used [2]. However, in the presence of a reasonable level of noise, the estimation variance of the LSYW method is still large. In order to overcome this problem, in [5], Davila has proposed a signal/noise subspace YW (SSYW) method by introducing a noise compensation in the LOYW method. In [6], by deriving a method of removing the noise-induced bias from the standard least-squares (LS) estimator, Zheng has proposed a method known as the improved least-squares fast-converging (ILSF) algorithm. Both the SSYW and ILSF methods are computationally fast

and provide estimation results that are quite satisfactory for low levels of signal-to-noise ratio (SNR).

Identifying AR systems from cepstral coefficients has been attempted only by a few researchers [7, 8]. In [7], a homomorphic LPC (HLPC) method has been proposed for both the white noise and periodic impulse-train excitations in a noise-free environment. The HLPC method cannot guarantee the stability of the estimated AR model. The ramp-cepstrum method proposed in [8] overcomes this problem even in a noisy environment. The method employs conventional cepstrum of a correlation function to formulate a ramp-cepstrum for the estimation of the AR parameters. Since conventional cepstrum is based on Fourier analysis, even for applications dealing with real data it involves complex computation and phase unwrapping operation during its implementation via the discrete Fourier transform (DFT) and inverse DFT (IDFT). In comparison to the DFT, the discrete cosine transform (DCT) is much better in many applications dealing with real signals, for example speech enhancement and speech recognition [1, 9], since it avoids complex computations. Also DCT requires relatively less number of coefficients to represent the signal/image data as compared to DFT. Moreover, it uses a very simple algorithm for phase unwrapping. Nevertheless, the DCT has rarely been employed for system identification problems [10]. In [10], the real spectrum computed by the DCT is employed to obtain an AR model describing the squared Hilbert temporal envelope of a sequence. In this paper, motivated by the advantageous features of DCT, we develop an AR system identification technique in the cepstral domain where DCT rather than DFT is employed. To this end, unlike the conventional cepstrum determined via Fourier and inverse Fourier transforms, a cosine cepstrum is first formally defined through cosine and inverse cosine transforms, and then utilized to develop a theory for the AR parameter estimation.

The objective of this paper is to develop an effective cosine-cepstrum-based methodology for the identification of AR systems from very heavily noise-corrupted samples of the output observations. The main idea of the proposed methodology is to achieve the above stated goal by having a transformed version of the corrupted signal for model-fitting so that the process of transformation itself is noise-robust, and by developing a corresponding target model for the purpose of fitting. In the proposed technique, the noise-robust approach of obtaining the transformed signal is to use the ramp cosine cepstrum (RCC) of one-sided autocorrelation function (OSACF) of an AR signal for both white noise and periodic impulse-train excitations. With this transformation for the signal, we are able to develop the corresponding target model, referred to as the RCC model, for the estimation of the system parameters. The motivation behind using the OSACF for the cosine cepstrum computation is to reduce the effect of the noise. Unlike conventional methods, we deal with both white noise and periodic impulse-train excitations. By employing the RCC model, a residue-based least-squares (RBLS) optimization scheme is presented for the estimation of the AR parameters. For the purpose of implementation, the DCT, which is

capable of handling the phase unwrapping problem and offers computational advantages over the DFT, is employed in the proposed method. The proposed method is tested for the estimation of the AR parameters of different synthetic AR systems and also for the identification of a human vocal tract system using natural speech signals.

The paper is organized as follows. In Section 2, the problem of AR system identification in the presence of noise is formulated in the cepstral domain. In Section 3, first, a ramp cosine cepstrum model based on a one-sided ACF of an AR signal for the two types of input excitations is derived and then the DCT is employed for the realization of the derived model. Section 4 presents a residue-based least-squares optimization scheme for the AR parameter estimation using the proposed ramp cosine cepstrum model under noisy conditions. The performance of the proposed method is demonstrated in Section 5 through extensive computer simulations for both synthetic and natural speech signals. Finally, in Section 6, salient features of the proposed algorithm are summarized with some concluding remarks.

2. Problem Statement

The input-output relationship of a real causal stable linear time-invariant autoregressive (AR) system can be described as

$$x(n) = -\sum_{k=1}^{M} a_k x(n-k) + w(n), \tag{1}$$

where $w(n)$ and $x(n)$ are, respectively, the excitation and the response of the AR system, $\{a_k\}$ the AR parameters to be estimated, and M the system order assumed to be known in this paper. Note that when the system order is unknown, different standard techniques, available in the literature [2], can be employed to estimate the order. The system output in (1) can be considered as a convolution of the input $w(n)$ and the impulse-response $h(n)$ of the system, represented as

$$x(n) = h(n) * w(n). \tag{2}$$

The transfer function of the AR(M) system described by (1) can be written as

$$H(z) = \frac{1}{A(z)} = \frac{1}{\prod_{k=1}^{M}(1 - p_k z^{-1})}, \tag{3}$$

where $A(z) = 1 + \sum_{k=1}^{M} a_k z^{-k}$ is the AR polynomial and $p_k = r_k e^{j\omega_k}$ represents the kth pole with a magnitude r_k and angle ω_k. In most of the system identification problems, $w(n)$ is modeled to be a stationary zero-mean white Gaussian noise with an unknown variance σ_w^2. For some practical applications, such as speech signal processing, seismology, and communication, however, the excitation may have other forms [1, 11–13]. For example, in speech signal processing, a periodic impulse-train is often used as an excitation of the vocal tract system [1, 11, 13]. As such, in this paper, both the white Gaussian noise and the periodic impulse-train excitations are considered as input to the AR system.

Cepstrum analysis has become a very important tool in signal processing, especially in different speech processing applications. It has been proposed as a method for separating signals that have been combined through convolution [1, 11]. For an N-point real sequence $\{s(n)\}_{n=0}^{N-1}$, in general, the cepstrum of $s(n)$ can be defined as [9]

$$\gamma_s(n) = \mathcal{T}^{-1}[\ln[\mathcal{T}[s(n)]]], \qquad (4)$$

where $\mathcal{T}[\,\cdot\,]$ and $\mathcal{T}^{-1}[\,\cdot\,]$, respectively, represent a transform and its inverse operator. When $\mathcal{T}$ is a z-transform, for example, $\mathcal{T}[s(n)] = S(z) = |S(z)|e^{\angle S(z)}$, and the natural logarithm yields

$$\ln[S(z)] = \ln[|S(z)|] + j\angle S(z). \qquad (5)$$

Definition in (4) is valid provided $s(n)$ is deterministic. Since a numerical computation of (5) provides only the principal or wrapped phase, a phase unwrapping algorithm is necessary to restore the phase continuity [11, 14].

In the current system identification problem, the system response $x(n)$, as described in (2), is a convolution of the input and the impulse-response of the system. In such a situation, (2) can be expressed in the cepstral domain by applying (4), where $\mathcal{T}[\,\cdot\,]$ is either z-transform or Fourier transform, as

$$\gamma_x(n) = \gamma_h(n) + \gamma_w(n), \qquad (6)$$

where $\gamma_h(n)$ is the cepstrum of the impulse response and $\gamma_w(n)$ represents the cepstrum of one realization of the input signal. Utilizing such an advantage of homomorphic deconvolution, cepstrum domain methods have been proposed for system identification in [7, 15, 16]. For example, in [7], in order to estimate the AR parameters, a mean-squared error minimization involving (6) is used by employing the Cholesky decomposition. However, as mentioned in [7], the problem of this method is that the stability of the estimated AR model is not guaranteed. It is to be noted that all the cepstral domain methods mentioned above deal only with the noise-free environment.

In the presence of additive noise $v(n)$, the observed signal $y(n)$ is given by

$$y(n) = x(n) + v(n), \qquad (7)$$

where $v(n)$ is assumed to be a zero mean stationary process and is independent of $w(n)$. In [17], the behavior of the cepstral coefficients in the presence of additive noise has been investigated for the purpose of speech recognition by assuming that the noise spectrum can be obtained during the experiment, and it has been shown that the cepstral vector of noisy data can be expressed as the sum of the cepstral vector of its clean version and a scaled deviation vector. In our identification problem, however, we handle a more common and critical situation where only noisy observations are available. Given one realization of input

excitation and the observation noise, using the definition in (4), the complex cepstrum of $y(n)$ can be expressed as

$$\gamma_y(n) = \mathcal{T}^{-1}\{\ln[\mathcal{T}[x(n)]]\} + \mathcal{T}^{-1}\left\{\ln\left[1 + \frac{\mathcal{T}[v(n)]}{\mathcal{T}[x(n)]}\right]\right\}$$
$$= \gamma_x(n) + \gamma_u(n), \qquad (8)$$

where $\gamma_u(n)$ arises because of the noise. The term $\gamma_u(n)$ determines as to how the noise affects $\gamma_y(n)$ and it vanishes altogether in the absence of noise. In order to estimate the AR system parameters from $\gamma_y(n)$, the effect of $\gamma_u(n)$ has to be reduced. It is difficult to obtain an accurate estimate of $\gamma_x(n)$ from $\gamma_y(n)$, since the cepstrum decomposition techniques are very sensitive to the noise level [17, 18]. In this paper, in order to reduce the effect of noise in extracting the AR parameters, first, we avoid computing cepstrum directly from the noise-corrupted observations by using a one-sided ACF, and then develop a ramp cosine cepstrum (RCC) model for a model-fitting based least-squares optimization in the cepstral domain. Moreover, in the proposed method, the DCT, instead of the conventional DFT, is employed for computing the cepstrum so as to overcome the problem of phase unwrapping and to achieve computational savings in dealing with real signals.

3. Proposed Ramp Cosine Cepstrum (RCC) Model Based on One-Sided ACF

In the cepstral analysis, cepstral coefficients are, generally, computed from an observed signal or from an estimate of its nonparametric power spectral density (PSD) [2, 19]. In this section we propose to develop a ramp cosine cepstrum model utilizing a one-sided ACF (OSACF) of $x(n)$, which can be defined as

$$\psi_x(\tau) = \begin{cases} \phi_x(\tau), & \tau > 0 \\ 0.5\phi_x(\tau), & \tau = 0 \\ 0, & \tau < 0 \end{cases} \qquad (9)$$

where $\phi_x(\tau)$ is the conventional two-sided ACF of $x(n)$ which, in general, is estimated as [2, 13]

$$\phi_x(\tau) = \frac{1}{N}\sum_{n=0}^{N-1-|\tau|} x(n)x(n+|\tau|), \quad 0 \le |\tau| < N, \qquad (10)$$

where N is the data length. This equation provides an accurate estimate of $\phi_x(\tau)$ when N is sufficiently large. Some important properties of the OSACF of $x(n)$ relevant to the development of the proposed model can be summarized as follows.

(1) As $\phi_x(\tau)$ is a symmetric two-sided sequence, the corresponding OSACF $\psi_x(\tau)$ is related to $\phi_x(\tau)$ by

$$\phi_x(\tau) = \psi_x(\tau) + \psi_x(-\tau). \qquad (11)$$

(2) For a real signal $x(n)$, its OSACF $\psi_x(\tau)$ is also real.

(3) The function $\psi_x(\tau)$ retains the pole-preserving property of $\phi_x(\tau)$.

(4) The OSACF exhibits a higher noise immunity than the conventional ACF does [20]. Since the spectral envelope of the OSACF of noisy observations, in comparison to the conventional two-sided conventional ACF, strongly enhances the highest power frequency bands corresponding to the spectral peaks, a large attenuation of the noise components lying outside the enhanced frequency bands would occur.

Taking the z-transform of both sides of (11) results in

$$\Phi_x(z) = \Psi_x(z) + \Psi_x\left(\frac{1}{z}\right). \tag{12}$$

The Fourier domain representation of (11) is given by

$$\mathcal{F}[\phi_x(\tau)] = 2\mathfrak{Re}[\mathcal{F}[\psi_x(\tau)]], \tag{13}$$

where $\mathcal{F}[\cdot]$ represents the Fourier transform and the operator $\mathfrak{Re}[\cdot]$ gives the real part of a complex number. As we are interested to perform cepstrum domain computation with $\psi_x(\tau)$, the relation in (13) favors the use of the cosine transform, which is the real part of the Fourier transform. The cosine transform, denoted as $\mathcal{F}_c[\cdot]$, of a real signal $\{\psi_x(n)\}_{n=0}^{N-1}$ can be written as

$$\Psi_x^c(\omega) = \mathcal{F}_c[\psi_x(n)] = \mathfrak{Re}[\mathcal{F}[\psi_x(n)]] = \sum_{n=0}^{N-1} \psi_x(n)\cos\omega n. \tag{14}$$

From (4) and (14), one can define the cosine cepstrum of a real signal $\{\psi_x(n)\}_{n=0}^{N-1}$ as

$$c_{\psi_x}(n) = \mathcal{F}_c^{-1}[\ln[\mathcal{F}_c[\psi_x(n)]]], \tag{15}$$

where $\mathcal{F}_c^{-1}[\cdot]$ denotes the inverse operator for the cosine transform, that is, for a given frequency domain spectrum $\Psi_x^c(\omega)$, the inverse cosine transform can be defined as

$$\mathcal{F}_c^{-1}[\Psi_x^c(\omega)] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Psi_x^c(\omega)\cos\omega n\,d\omega. \tag{16}$$

In the following, we will develop a ramp cosine cepstrum model for the estimation of the AR parameters under the white Gaussian noise and periodic impulse-train excitations. To this end, we first show that the cosine cepstrum $c_{\psi_x}(n)$ can be expressed in terms of the system poles. Using (13) and (14), $c_{\psi_x}(n)$ in (15) can be expressed as

$$c_{\psi_x}(n) = \mathcal{F}_c^{-1}[\ln[\mathcal{F}[\phi_x(n)]]] + \mathcal{F}_c^{-1}\left[\ln\left[\frac{1}{2}\right]\right]. \tag{17}$$

Here, $\mathcal{F}[\phi_x(n)] = \Phi_x(\omega)$ is by definition the PSD of the real signal $x(n)$, and it can be shown that $\Phi_x(\omega)$ is real, even, and nonnegative. From (2), the PSD of the output $x(n)$ for the linear time-invariant system with the transfer function $H(z)$ given by (3) can be expressed as

$$\Phi_x(\omega) = H(\omega)H(-\omega)\Phi_w(\omega) = |H(\omega)|^2\Phi_w(\omega), \tag{18}$$

where $\Phi_w(\omega)$ is the PSD of the input signal. Using (18), $c_{\psi_x}(n)$ in (17) can be written as

$$c_{\psi_x}(n) = \mathcal{F}_c^{-1}[\ln[H(\omega)]] + \mathcal{F}_c^{-1}[\ln[H(-\omega)]]$$
$$+ \mathcal{F}_c^{-1}[\ln[\Phi_w(\omega)]] + \mathcal{F}_c^{-1}\left[\ln\left[\frac{1}{2}\right]\right]. \tag{19}$$

It is observed from (19) that the effect of input excitation $w(n)$ has been made additive by using the homomorphic deconvolution. Now, we consider each of the four terms in (19) individually. From (3), $\ln[H(z)]$ can be expanded as

$$\ln[H(z)] = -\sum_{i=1}^{M}\ln(1 - p_iz^{-1}) = \sum_{i=1}^{M}\sum_{n=1}^{\infty}\frac{p_i^n}{n}z^{-n}, \tag{20}$$

where $|z| > |p_i|$. Using (16), the inverse cosine transform of $\ln[\mathcal{F}[h(n)]]$, with $h(n)$ being real and minimum phase, can be calculated by

$$\mathcal{F}_c^{-1}[\ln[H(\omega)]] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\sum_{i=1}^{M}\sum_{m=1}^{\infty}\frac{p_i^m}{m}e^{-j\omega m}\right]\cos\omega n\,d\omega$$
$$= \frac{1}{2\pi}\sum_{i=1}^{M}\sum_{m=1}^{\infty}\frac{p_i^m}{m}\int_{-\pi}^{\pi}e^{-j\omega m}\cos\omega n\,d\omega. \tag{21}$$

Noting that

$$\int_{-\pi}^{\pi}e^{-j\omega m}\cos\omega n\,d\omega = \begin{cases} \pi, & m = n \\ 0, & m \neq n, \end{cases} \tag{22}$$

we have

$$\mathcal{F}_c^{-1}[\ln[H(\omega)]] = \frac{1}{2}\sum_{i=1}^{M}\frac{p_i^n}{n}, \quad n > 0. \tag{23}$$

Similarly, the inverse cosine transform of $\ln[H(-\omega)]$ can be obtained as

$$\mathcal{F}_c^{-1}[\ln[H(-\omega)]] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\sum_{i=1}^{M}\sum_{m=1}^{\infty}\frac{p_i^m}{m}e^{j\omega m}\right]\cos\omega n\,d\omega$$
$$= \frac{1}{2}\sum_{i=1}^{M}\frac{p_i^n}{n}, \quad n > 0. \tag{24}$$

It is observed from (16) that for a constant value of $\Psi_x^c(\omega)$, $\mathcal{F}_c^{-1}[\Psi_x^c(\omega)] = 0$ for all $n > 0$. Thus for $n > 0$, the last term on the right side of (19) vanishes. Let us now consider the remaining third term of (19) that depends on the characteristics of the input excitation $w(n)$. In the following section we consider separately the white Gaussian noise and a periodic impulse-train as an input excitation.

3.1. White Noise Excitation. For a zero mean white Gaussian noise with a variance σ_w^2, $\Phi_w(\omega) = \sigma_w^2$. Thus, the third term on the right side of (19) reduces to

$$\mathcal{F}_c^{-1}[\ln[\Phi_w(\omega)]] = \mathcal{F}_c^{-1}[\ln[\sigma_w^2]] = 0, \quad n > 0. \tag{25}$$

Hence, for the white noise excitation, the cosine cepstrum $c_{\psi_x}(n)$ in (19) can finally be expressed as

$$c_{\psi_x}(n) = \sum_{i=1}^{M} \frac{p_i^n}{n}, \quad n > 0. \qquad (26)$$

It can be observed from this equation that $c_{\psi_x}(n)$ decays rapidly with increasing n, thus making it difficult to use $c_{\psi_x}(n)$ for the estimation of the system poles. In order to overcome this problem, we propose an easy-to-handle ramp cosine cepstrum (RCC) for the OSACF of $x(n)$, defined as

$$\chi_x(n) = n c_{\psi_x}(n) = \sum_{i=1}^{M} p_i^n, \quad n > 0. \qquad (27)$$

Since the poles in a system could appear as real or as complex conjugate pair, (27) can be rewritten as

$$\chi_x(n) = \sum_{i=1}^{\kappa} \alpha(\omega_i) r_i^n \cos(\omega_i n), \quad n > 0, \qquad (28)$$

where κ is the number of real poles plus the number of complex conjugate pole pairs, r_i and ω_i are, respectively, the magnitude and the argument of p_i. In (28), $\alpha(\omega_i)$ is introduced to distinguish real and complex poles and is given by

$$\alpha(\omega_i) = \begin{cases} 1, & \omega_i = 0 \text{ or } \omega_i = \pi, \\ 2, & 0 < \omega_i < \pi. \end{cases} \qquad (29)$$

The model given by (28) is termed as the AR ramp cosine cepstrum (RCC) model for the OSACF of $x(n)$. This model will be used in the next section to formulate an objective function for the least-squares fitting problem in a noisy environment.

3.2. Periodic Impulse-Train Excitation. In the derivation of the RCC model with the white noise excitation, it was observed that the term containing the effect of white noise excitation becomes zero for $n > 0$, since the PSD of the input $w(n)$ is a constant. However, the situation is more complicated in the case of a periodic impulse-train excitation $w_i(n)$ where the corresponding PSD is no longer a constant. Next, we analyze the effect of the third term $\mathcal{F}_c^{-1}[\ln[\Phi_{w_i}(\omega)]]$ of (19), which is now denoted as $\hat{c}_{\phi_{wi}}(n)$, on $c_{\phi_x}(n)$.

A periodic impulse-train excitation $\{w_i(n)\}_{n=0}^{N-1}$ with a given period T can be expressed as [13]

$$w_i(n) = \sum_{k=0}^{\mu-1} \delta(n - kT), \qquad (30)$$

where $\mu = \lceil N/T \rceil$, $\lceil \cdot \rceil$ denoting the ceiling operator, is the total number of impulses within the finite duration of excitation. Using (10), an estimate of the ACF of $w_i(n)$ is obtained as

$$\phi_{w_i}(\tau) = \frac{1}{N} \sum_{l=0}^{\mu-1} (\mu - l)\delta(|\tau| - lT), \quad 0 \le |\tau| < N. \qquad (31)$$

It is observed from (31) that $\phi_{w_i}(\tau)$ decays with increasing values of τ and has nonzero values at $\tau = 0$ and at integer multiples of T for the case of finite data operation with $0 \le |\tau| < N$. Thus, $\phi_{w_i}(\tau)$ can be expressed alternately as

$$\phi_{w_i}(\tau) = \begin{cases} f\left(\dfrac{\tau}{T}\right), & |\tau| = 0, T, 2T, \ldots, (\mu-1)T, \\ 0, & \text{otherwise,} \end{cases} \qquad (32)$$

where

$$f(\tau) = \begin{cases} \dfrac{\mu - |\tau|}{N}, & |\tau| \le \mu - 1, \\ 0, & \text{otherwise.} \end{cases} \qquad (33)$$

Note that $f(\tau)$ is an even symmetric triangular sequence and from (32) and (33), it is evident that $f(\tau)$ can be obtained by down-sampling $\phi_{w_i}(\tau)$ with a factor T. Thus the z transform of $\phi_{w_i}(\tau)$ can be expressed as

$$\Phi_{w_i}(z) = F\left(z^T\right), \qquad (34)$$

where $F(z)$ is the z transform of $f(n)$ and the sequence $Nf(n)$ can be generated through a convolution between a rectangular pulse train of width μ and its time reversal sequence. An expression for $F(z)$ can be obtained as

$$F(z) = \frac{1}{N} \frac{(z^\mu - 1)^2}{z^{\mu-1}(z-1)^2}, \quad z \neq 0, 1. \qquad (35)$$

Based on the relation between $\phi_{w_i}(\tau)$ and $f(\tau)$, as described in (32), (33), and (34), it can be shown that

$$\hat{c}_{\phi_{wi}}(n) = \begin{cases} \hat{c}_f\left(\dfrac{n}{T}\right), & n = 0, T, 2T, \ldots, (\mu-1)T, \\ 0, & \text{otherwise,} \end{cases} \qquad (36)$$

where

$$\hat{c}_f(n) = \mathcal{F}_c^{-1}[\ln[F(\omega)]]. \qquad (37)$$

It is evident from (36) that $\hat{c}_{\phi_{wi}}(n)$ assumes nonzero values at $n = 0$ and at integral multiples of T for $n \ge 0$. Thus, the third term on the right side of (19) reduces to

$$\mathcal{F}_c^{-1}[\ln[\Phi_{w_i}(\omega)]] = 0, \quad 0 < n < T. \qquad (38)$$

Note that the RCC given by (27) for the white noise excitation can be modified for the impulse-train excitation as

$$\chi_x(n) = n c_{\psi_x}(n) = \sum_{i=1}^{M} p_i^n, \quad 0 < n < T. \qquad (39)$$

From (27) and (39), it is observed that the RCC model derived for the white noise excitation is also valid for the case of periodic impulse-train excitation when $0 < n < T$.

3.3. Computation of RCC Model via DCT/IDCT. The RCC model derived in the previous subsections is obtained from the cosine cepstrum of the OSACF of $x(n)$, where the logarithm operation is performed on the cosine transform of $\psi_x(m)$. As explained earlier, the difficulty in the complex cepstral analysis is the necessity to unwrap the phase to make it a continuous function of ω. A major advantage of using cosine transform lies in its binary phase information, i.e., 0 or π which, as shown later, can significantly simplify the phase unwrapping process. From the implementation point of view, different types of discrete cosine-transforms (DCTs) can be employed. It is known that the DCT is far superior to the DFT for the transformation of real signals. For a real signal, DFT gives complex spectrum and leaves nearly one-half of data unused. In contrast, the DCT generates real spectrum of real signals and thereby makes the computation of redundant data unnecessary. Being a real function, the DCT offers an added advantage that it requires only a simple phase unwrapping algorithm. Also, as the DCT is derived from the DFT, all the desirable properties of DFT are preserved, and fast algorithms for its computation exist. As a result, using a DCT and inverse DCT (IDCT) pair, a complex-cepstrum corresponding to (15) can be implemented as follows

$$\chi_x(n) = nc_{\psi_x}(n), \quad n > 0,$$
$$c_{\psi_x}(n) = \text{IDCT}[\ln(\text{DCT}[\psi_x(n)])], \quad n > 0. \tag{40}$$

For a real sequence $\psi_x(n)$ with $n = 0, 1, \ldots, N-1$, the most commonly used DCT-IDCT pair is defined as

$$\text{DCT}[\psi_x(n)] = \Psi_x(k) = \rho(k) \sum_{n=0}^{N-1} \psi_x(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right),$$
$$k = 0, 1, \ldots, N-1, \tag{41}$$

$$\text{IDCT}[\Psi_x(k)] = \psi_x(n) = \sum_{k=0}^{N-1} \rho(k)\Psi_x(k) \cos\left(\frac{(2n+1)k\pi}{2N}\right),$$
$$n = 0, 1, \cdots, N-1, \tag{42}$$

where $\rho(k)$ is a normalization coefficient defined as

$$\rho(k) = \begin{cases} \sqrt{\dfrac{1}{N}}, & \text{for } k = 0, \\ \sqrt{\dfrac{2}{N}}, & \text{for } k = 1, 2, \ldots, N-1. \end{cases} \tag{43}$$

Since the bases of the cosine transform are real functions, the principal phases of DCT coefficients can only be 0 or π. Accordingly, we can represent the phase as $\exp(-j\pi)$ when the cosine transform is negative sign and as $\exp(-j0)$ when it is positive. With this representation, the logarithm operation in (40) can be easily carried out and (40) can be expressed as

$$\varsigma_x(n) = \Re e[\text{IDCT}[\ln|\Psi_x(k)| + j\pi\xi]], \quad n > 0, \tag{44}$$

where

$$\xi = \begin{cases} 0, & \text{if } \Psi_x(k) \geq 0, \\ -1, & \text{if } \Psi_x(k) < 0. \end{cases} \tag{45}$$

Thus, this representation clearly supports a simple phase unwrapping. On the other hand, in the case of using DFT for the computation of cepstrum, complicated phase unwrapping algorithms as proposed in literature [11, 14] need to be used, since the phase in this case has no longer binary values.

4. The RCC Model-Based Parameter Estimation

4.1. Effect of Additive Noise. In the presence of noise, the observed signal gets heavily corrupted especially when the signal-to-noise ratio (SNR) is very low. In this paper, a more general noisy environment is considered where it is assumed that the noise variance is unknown and noise-only data is not available. In Section 2, the effect of noise on cepstral coefficients has been described for the case when cepstrum is computed in the signal domain. It is well-known that the autocorrelation of a noisy signal offers more noise-robustness in comparison to the noisy signal itself [20]. Thus, the RCC model that we have developed based on the OSACF $\psi_x(\tau)$ of noise-free signal can be used as a target function even when RCC is computed based on the OSACF of the noisy observation of the signal. In what follows, our objective is to investigate the effect of the noise on the RCC computed from noisy observations. In the presence of an additive noise $v(n)$, the ACF of the noisy observation $y(n)$ can be expressed as

$$\phi_y(\tau) = \phi_x(\tau) + \phi_n(\tau), \tag{46}$$

where

$$\phi_n(\tau) = \phi_v(\tau) + \phi_{xv}(\tau) + \phi_{vx}(\tau). \tag{47}$$

Here, $\phi_v(\tau)$ is the ACF of noise $v(n)$, and $\phi_{xv}(\tau)$ and $\phi_{vx}(\tau)$ are crosscorrelation terms. Equation (46) is valid for both the estimated and the theoretical ACFs. It can be observed that $\phi_n(\tau)$ corrupts $\phi_x(\tau)$ in an additive fashion like the signal. The effect of $\phi_n(\tau)$ cannot be neglected, especially when the SNR is very low. Note that the effect of crosscorrelation terms on $\psi_x(\tau)$ is negligible when $v(n)$ and $w(n)$ are assumed to be uncorrelated. However, at a very low SNR, this is not so when the length of the observed data is finite. Even for an uncorrelated additive white Gaussian noise, all the lags of the noisy ACF are corrupted at a very low SNR. Under such a noisy condition, the conventional correlation based methods employing directly $\phi_y(\tau)$ cannot provide a good estimation performance. This motivates us to switch to the cepstral domain where the logarithmic smoothing would help in preserving the RCC model under heavy noisy

conditions. The OSACF $\psi_y(\tau)$ of noisy observations $y(n)$ can be obtained as

$$\psi_y(\tau) = \begin{cases} \phi_y(\tau), & \tau > 0, \\ 0.5\phi_y(\tau), & \tau = 0, \\ 0, & \tau < 0. \end{cases} \quad (48)$$

From (46) and (48), the OSACF of $y(n)$ can be written as

$$\psi_y(\tau) = \psi_x(\tau) + \psi_n(\tau), \quad (49)$$

where $\psi_n(\tau)$ indicates the effect of noise on $\psi_y(\tau)$ and it can be expressed in a form similar to that of $\psi_x(\tau)$ given by (9). Thus, in the presence of noise, the cosine cepstrum of $\psi_y(\tau)$ can be expressed as

$$c_{\psi_y}(n) = \mathcal{F}_c^{-1}\Big[\ln\big[\mathcal{F}_c\big[\psi_y(n)\big]\big]\Big] = c_{\psi_x}(n) + c_{\psi_n}(n), \quad n > 0, \quad (50)$$

where

$$c_{\psi_n}(n) = \mathcal{F}_c^{-1}\left\{\ln\left[1 + \frac{\mathcal{F}_c[\psi_n(n)]}{\mathcal{F}_c[\psi_x(n)]}\right]\right\}. \quad (51)$$

Therefore, the ramp cosine cepstrum of $\psi_y(\tau)$ can be expressed as

$$\chi_y(n) = \chi_x(n) + \chi_\varepsilon(n), \quad n > 0. \quad (52)$$

Here, the term $\chi_\varepsilon(n)$ arises because of the noise. Like $\gamma_u(n)$ in (8), $\chi_\varepsilon(n)$ would vanish in the absence of noise. Now, the RCC model derived in Section 3 can be used in (52) for a ramp cosine cepstral model fitting to minimize the error between $\chi_y(n)$ and $\chi_x(n)$. By this approach the RCC model parameters, and thus the AR parameters are estimated.

Since, in the presence of additive white Gaussian noise, the zero lag of the noisy ACF $\phi_y(n)$ is most severely corrupted in comparison to other lags, if the zero lag is kept as it is during the computation of the RCC of the OSACF, it may result in a more erroneous value of RCC. On the other hand, excluding the zero lag, although it may reduce the effect of noise, would remove the average power of the observed data $y(n)$. Since $\phi_y(0) > |\phi_y(\tau)|$ for $\tau \neq 0$, we replace $\phi_y(0)$ by $\eta\phi_y(0)$ with $\{|\phi_y(1)|/\phi_y(0)\} \leq \eta < 1$ in order to reduce the effect of noise. This is suitable especially for a difficult situation where noise variance and/or noise-only data are not available. The process can efficiently suppress the level of $c_{\psi_n}(n)$ while leaving the shape of $c_{\psi_y}(n)$ similar to that of $c_{\psi_x}(n)$.

4.2. Ramp Cosine Cepstral Fitting: Residue-Based Least-Squares Optimization. As discussed in the previous subsection following (52) that a ramp cosine cepstral fitting approach can be developed to determine the RCC model parameters from the RCC of the OSACF of noisy observations. We now propose a residue-based least-squares (RBLS) fitting scheme to estimate the model parameters in (28) and (39). Then, the AR parameters can be obtained from the RCC model parameters $\{r_i\}$ and $\{\omega_i\}$. Each of the κ

component terms in (28) contains a pair (r_i, ω_i). In order to estimate each of the κ such pairs, N_c values of $\chi_y(n)$ are used, where $N_c < T$ for the periodic impulse-train excitation. The objective function to determine the values of one pair (r_i, ω_i) is defined as the total squared error between the $(l-1)$th residual function $\mathfrak{R}_{l-1}(n)$ and the lth component of the RCC model, that is

$$J_l = \sum_{n=1}^{N_c} \left| \mathfrak{R}_{l-1}(n) - \alpha(\omega_l)r_l^n\cos(\omega_l n) \right|^2, \quad l = 1, 2, \ldots, \kappa, \quad (53)$$

where the residual function is updated as follows

$$\mathfrak{R}_0(n) = \chi_y(n),$$

$$\mathfrak{R}_l(n) = \mathfrak{R}_{l-1}(n) - \alpha(\omega_l)r_l^n\cos(\omega_l n), \quad l = 1, \ldots, \kappa - 1. \quad (54)$$

Note that $\{r_l\}$ and $\{\omega_l\}$ are independent variables and α depends on $\{\omega_l\}$ as seen from (29). We would like to find the optimal solution for $\{r_l\}$ and $\{\omega_l\}$ by a search algorithm based on the computation of (53) and (54). In order to reduce the computational burden, a two-step search algorithm is adopted. In the first step, a coarse-search based on the DCT spectrum of the OSACF of the observed data is employed to find out the initial estimate of $\{\omega_l\}$ and $\{r_l\}$, $l = 1, 2, \ldots \kappa$. In the second step, a fine-search is carried out around each initially estimated pair of $\{\omega_l\}$ and $\{r_l\}$ to obtain a more accurate estimate. In the fine-search, a neighborhood centered at each initial estimate of $\{r_l\}$ and $\{\omega_l\}$ is searched with a prescribed search resolution in a bounded region. A pair of $\{r_l'\}$ and $\{\omega_l'\}$ that globally minimizes J_l is selected as the estimate of a desired pole. It can be observed from (54) that, in order to determine the lth residual function $\mathfrak{R}_l(n)$, the computed values of $\{r_l\}$ and $\{\omega_l\}$ are utilized. Proceeding in this manner, the AR parameters can be determined using (3) once all the M poles have been estimated. In the proposed search scheme, restricting the search range of r_l within the stable region inherently guarantees the stability of the estimated AR system. Another advantage of the RBLS scheme is that in each fine-search, instead of the entire RCC model with all κ constituent terms, only one such term is estimated in (53), and in this fashion each term of the RCC model is sequentially obtained. This is done with a view to convert a multivariable optimization problem into a set of two-variable optimization scheme which makes the problem much simpler.

5. Simulation Results

In this section, extensive simulations are carried out in order to demonstrate the effectiveness of the proposed technique in identifying the AR systems in the presence of noise. We investigate the identification performance for synthetic AR signals as well as natural speech signals corrupted by additive noise. The estimation performance of the proposed method in terms of the accuracy and consistency of the estimated parameters is obtained and compared with that

of the existing methods including the improved least-squares fast-converging (ILSF) method [6], the signal/noise subspace Yule-Walker (SSYW) method [5], and the modified least-squares Yule-Walker (MLSYW) method [2].

5.1. Results on Synthetic AR Systems

5.1.1. White Noise Excitation. A noisy signal is generated according to (1) and (7) with $N = 4,000$ and $\sigma_w^2 = 1$, where the variance of the white Gaussian noise σ_v^2 is appropriately set based on a specified level of SNR defined as

$$\text{SNR} = 10 \log_{10} \frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} v(n)^2} \text{ dB.} \qquad (55)$$

From the noisy observations, first, the OSACF $\psi_y(\tau)$ is computed using (48) and (10). Note that for the purpose of implementing the cosine cepstrum, generally the continuous frequency ω is sampled as $\omega = 2\pi k/M$ for $k = 0, 1, \ldots, K-1$ resulting in a K-point Discrete Cosine Transform (DCT). According to the description provided in Section 3.3 for the noise-free observations, DCT-IDCT-based ramp-cosine cepstrum (RCC) is computed using $\psi_y(\tau)$. The RCC model parameters are then determined using the residue-based least-squares optimization technique introduced in Section 4.2. In the proposed optimization scheme, the search range for r_l is chosen in the range $[0.5, 0.99]$, that allows the identification of systems even with a very fast decaying autocorrelations. The initial estimates of ω_l are obtained from the location of the peaks of the smoothed DCT of the OSACF of $y(n)$. The search range for ω_l is in a range of 0.1π chosen symmetrically around the neighborhood of the initial estimates. Search resolutions of $\Delta r = 0.01$ and $\Delta \omega = 0.01\pi$ are used for r_l and ω_l, respectively. It has been experimentally found that, in order to obtain a better estimate of the M unknown AR coefficients, the number of RCC samples to be considered in the model-fitting operation should be higher than M. In our experiment, the number of RCC samples is taken as $N_c = 10M$.

As discussed in Section 4.1, in order to reduce the effect of the most corrupted zero lag on the OSACF of the noisy observations, the value of η is chosen as $|\phi_y(1)|/\phi_y(0)$. Several experiments, each consisting of $N_T = 100$ independent trials, are conducted to find the means and variances of the estimated AR parameters under noisy observations in which the SNR varies from -5 dB to 15 dB at steps of 2.5 dB. The performance measurement criteria considered in our simulation study are (1) the mean of estimated parameters, (2) the standard deviation from the mean (SDM), (3) the standard deviation from the given value, that is, the true value (SDT), and (4) the average sum-squared error (ASSE) given by

$$\text{ASSE} = \frac{1}{N_T M} \sum_{m=1}^{N_T} \sum_{k=1}^{M} [\hat{a}_k(m) - a_k]^2, \qquad (56)$$

where $\hat{a}_k(m)$ represents the estimated parameter at the mth trial and a_k the corresponding true value of the parameter.
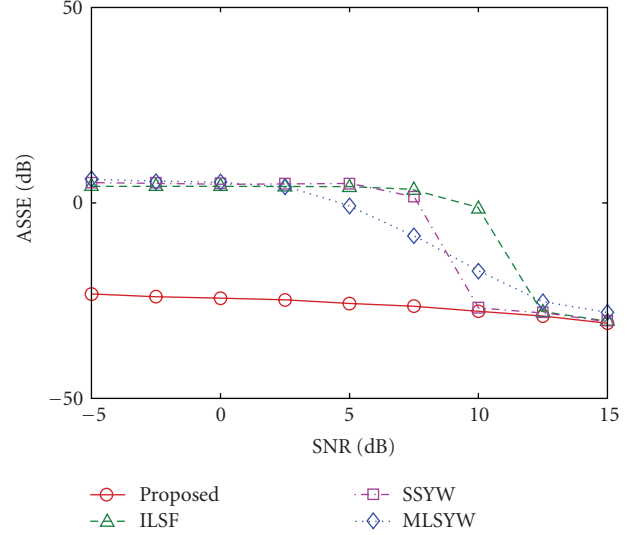

FIGURE 1: Effect of noise level on the ASSE for a white noise-excited system.

Different AR systems are investigated in order to cover a wide range of possible locations of poles, their numbers and types (i.e., real or complex conjugate). Tables 1 and 2 show the estimation results for the AR(3) and AR(4) systems at an SNR level of -5 dB, respectively. The AR(3) system with $A(z) = 1 - 2.6770z^{-1} + 2.5894z^{-2} - 0.8970z^{-3}$ contains a real pole and a pair of complex conjugate pole, and the AR(4) system with $A(z) = 1 + 0.4998z^{-1} - 0.01z^{-2} - 0.7853z^{-3} - 0.5999z^{-4}$ contains two real poles and a pair of complex conjugate poles. As the real and complex types of poles exhibit quite different behaviors, in our experiments various combinations of real and complex poles are considered to show the capability of the proposed algorithm in dealing with real life situations. In each table, the second column lists the true values of the AR parameters and the remaining four columns list the estimated values of corresponding parameters obtained from the proposed and the three other methods. The values for the SDM and SDT corresponding to estimated AR coefficients are also given below the estimated parameter value. The last row of each table provides the ASSE measure in dB. Table 1 shows that at SNR $= -5$ dB, when the other methods fail to identify the system, the proposed method successfully estimates the parameters quite accurately. It is seen from Table 2, although some of the other methods provide an acceptable performance, the estimation accuracy achieved by the proposed method is much higher. It is seen from these tables that the proposed method exhibits a superior estimation performance with respect to all the four performance indices at such a low level of SNR. Very small values of SDM and SDT obtained from the proposed technique indicate a high degree of estimation consistency and accuracy.

Figure 1 shows the ASSE values as a function of SNR levels for the AR(3) system obtained by each of the four methods with the true parameters as specified in Table 1. It is observed from Figure 1 that the ILSF and the SSYW methods

TABLE 1: Estimated parameters at SNR $= -5$ dB for AR(3) system with white noise excitation.

True parameters		Estimated parameters			
		Proposed method	ILSF method	SSYW method	MLSYW method
a_1	-2.6770	-2.6658	-1.3437	-0.9891	-1.0753
		(± 0.0349)	(± 0.5077)	(± 0.6716)	(± 0.1074)
		(± 0.0367)	(± 1.4267)	(± 1.8166)	(± 1.6053)
a_2	2.5894	2.5517	0.4403	-0.1136	0.1118
		(± 0.0760)	(± 1.0037)	(± 0.7117)	(± 0.1799)
		(± 0.0859)	(± 2.3719)	(± 2.5760)	(± 2.7071)
a_3	-0.8970	-0.8763	0.0868	0.2214	0.5107
		(± 0.0439)	(± 0.6053)	(± 0.3641)	(± 0.1013)
		(± 0.0496)	(± 1.1551)	(± 1.1762)	(± 1.4114)
ASSE (dB)		-24.93	4.25	5.13	6.05

TABLE 2: Estimated parameters at SNR $= -5$ dB for AR(4) system with white noise excitation.

True parameters		Estimated parameters			
		Proposed method	ILSF method	SSYW method	MLSYW method
a_1	0.4998	0.5042	0.3655	0.3830	1.0445
		(± 0.0289)	(± 0.2595)	(± 0.3086)	(± 0.0923)
		(± 0.0293)	(± 0.2922)	(± 1.6859)	(± 1.2579)
a_2	-0.0100	-0.0283	-0.0066	0.0040	0.0452
		(± 0.0219)	(± 0.0600)	(± 0.0651)	(± 0.0704)
		(± 0.0285)	(± 0.0601)	(± 0.0672)	(± 0.0747)
a_3	-0.7853	-0.7580	-0.7759	-0.8221	-0.7559
		(± 0.0507)	(± 0.0893)	(± 0.0857)	(± 0.0956)
		(± 0.0665)	(± 0.0899)	(± 0.0882)	(± 0.0972)
a_4	-0.5999	-0.5648	-0.4597	-0.4211	-0.3229
		(± 0.0374)	(± 0.2732)	(± 0.2874)	(± 0.3113)
		(± 0.0513)	(± 0.3071)	(± 0.2982)	(± 0.3257)
ASSE (dB)		-24.95	-13.27	-12.71	-9.13

give estimation accuracy comparable to that provided by the proposed method for SNR levels above 10 dB. However, the proposed method performs significantly better for levels of SNR as low as -5 dB.

Figure 2 depicts the superimposed plots of the estimated poles from 20 independent realizations obtained by the four methods at SNR $= -5$ dB along with their true locations for an AR(5) system with parameters $\{a_k\} = \{1, -3.2229, 5.2862, -5.0095, 2.7875, -0.7362\}$. Clearly, the estimated values obtained using the proposed method in comparison to that achieved by the other methods are much less scattered around the true values indicating a very high estimation accuracy. Similar to AR(3), AR(4), and AR(5) systems described above, the performance of the proposed method has been investigated for a number of other AR systems with different orders. As an illustration of the effectiveness of the proposed RCC method with larger model orders, an AR(12) system is considered with parameters $\{a_k\} = \{1, -2.1953, 3.7702, -5.7045, 7.9177, -9.0049, 9.2872, -8.8448, 7.5863, -5.3168, 3.4542, -1.9537, 0.8162\}$. In Figure 3, the superimposed plots of the estimated poles of the AR(12)

system obtained by the four methods at SNR $= -5$ dB along with their true locations are shown. Similar to Figure 2 that portrays the estimation accuracy of the proposed method for the case of AR(5) system, Figure 3 clearly exhibits the effectiveness of the proposed method in estimating poles of the high-order AR system. As expected, the estimation accuracy of this large order AR system is somewhat reduced, but the performance of the proposed RCC method still remains considerably superior to that provided by the other techniques.

5.1.2. Impulse-Train Excitation. We now consider the problem of AR system identification with periodic impulse-train excitations of different periods for various levels of noise. An impulse-train is generated using (30) with a known value of T. We choose the number of RCC samples less than T; thus, $N_c = \min(T - 1, 10M)$. A noisy AR signal is generated according to (1) and (7) with $N = 4,000$. The simulations are carried out for $N_T = 100$ independent trials and the results averaged.
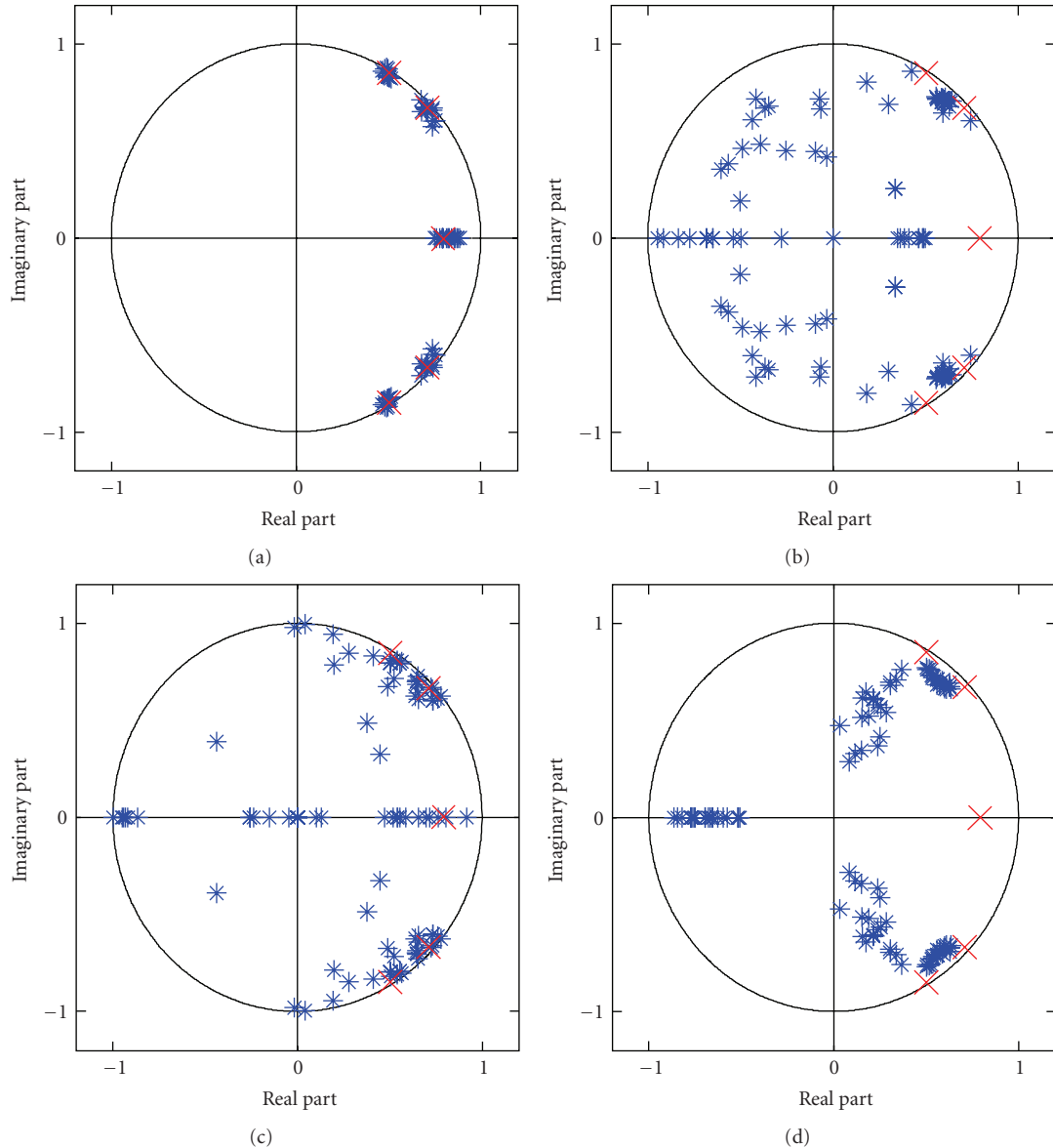
FIGURE 2: Superimposed pole plot of AR(5) system at SNR = −5 dB. ×: true poles and ∗: estimated poles. (a) Proposed, (b) ILSF, (c) SSYW, and (d) MLSYW method.

Tables 3 and 4 provide the estimation results for the impulse-train excited AR(3) and AR(4) systems with $T = 220$ at SNR = −5 dB, respectively. It is seen from these tables that the proposed method provides quite accurate estimation of the AR parameters with very small values of SDM and SDT, whereas the other methods are unable to identify the systems at SNR = −5 dB. Similar result is observed for the AR(5) system that was considered for the white noise excitation.

The ASSE resulting from using the various methods under the impulse-train excitation for the estimation of the same AR(3) system as the one considered for the white noise excitation is shown in Figure 4. It is seen from the figure that, the proposed RCC method provides a significantly better performance even at a very low SNR, whereas the

performance of other methods deteriorates at low levels of SNR.

It is to be mentioned that, we have also compared the proposed ramp cosine cepstrum (RCC) method with our ramp cepstrum (RC) method previously developed in [8] which employs conventional cepstrum of a correlation function via DFT and IDFT. It has been observed that the estimation performance of the RCC method is slightly better than our previous RC method at a very low SNR of around −5 dB, and remains comparable for other levels of SNR. Although the two methods exhibit quite a similar estimation performance, yet the RCC method based on the DCT-IDCT implementation offers significant computational advantages as opposed to the RC approach.
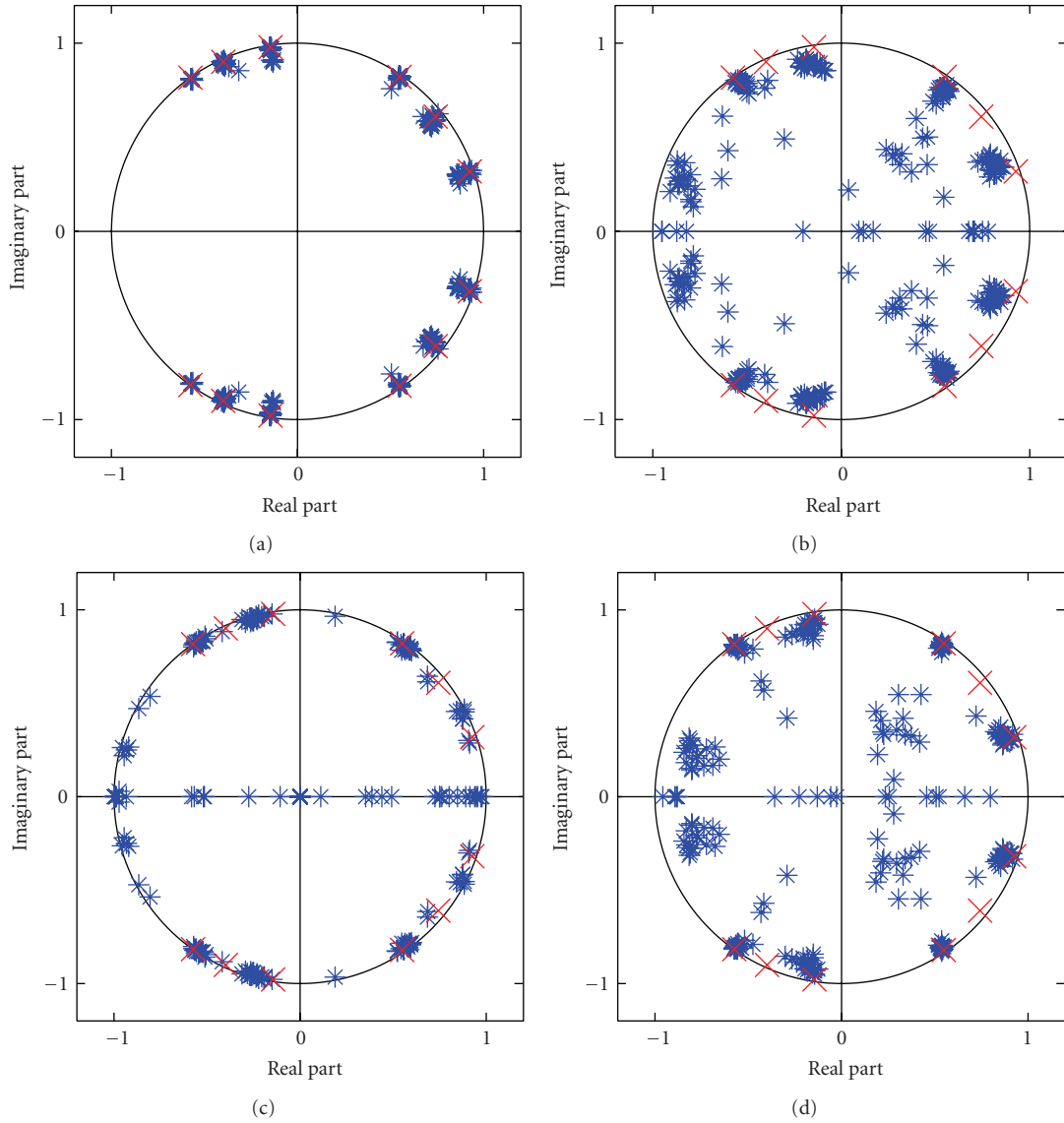
FIGURE 3: Superimposed pole plot of AR(12) system at SNR = −5 dB. ×: true poles and ∗: estimated poles. (a) Proposed, (b) ILSF, (c) SSYW, and (d) MLSYW method.

TABLE 3: Estimated parameters at SNR = −5 dB for AR(3) system with impulse train excitation.

	True parameters	Estimated parameters			
		Proposed method	ILSF method	SSYW method	MLSYW method
		−2.6816	−0.9615	−1.0776	−1.0588
a_1	−2.6770	(±0.0311)	(±1.0657)	(±0.6890)	(±0.0962)
		(±0.0290)	(±2.0196)	(±1.7414)	(±1.6211)
		2.5644	−0.2284	0.1767	−0.1307
a_2	2.5894	(±0.0702)	(±1.7218)	(±0.7101)	(±0.1728)
		(±0.0671)	(±3.3022)	(±2.5150)	(±2.7256)
		−0.8732	0.4773	0.2489	0.5269
a_3	−0.8970	(±0.0398)	(±0.8970)	(±0.3625)	(±0.0987)
		(±0.0387)	(±1.6411)	(±1.2019)	(±1.4273)
ASSE (dB)		−25.14	5.23	6.27	5.87

TABLE 4: Estimated parameters at SNR = −5 dB for AR(4) system with impulse train excitation.

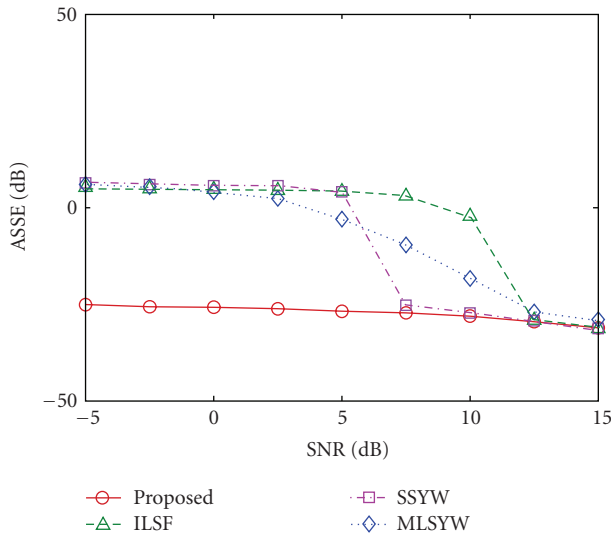| True parameters		Estimated parameters			
		Proposed method	ILSF method	SSYW method	MLSYW method
a_1	0.4998	0.4822 (± 0.0432) (± 0.0456)	0.3845 (± 0.2824) (± 0.2914)	0.3719 (± 0.3122) (± 0.3134)	0.1483 (± 0.4145) (± 0.4225)
a_2	−0.0100	−0.0591 (± 0.0501) (± 0.0540)	0.0151 (± 0.0705) (± 0.0743)	0.0247 (± 0.0607) (± 0.0699)	0.0608 (± 0.0615) (± 0.0938)
a_3	−0.7853	−0.7483 (± 0.0651) (± 0.0730)	−0.8134 (± 0.0602) (± 0.0664)	−0.8428 (± 0.0402) (± 0.0701)	−0.7973 (± 0.0387) (± 0.0407)
a_4	−0.5999	−0.5568 (± 0.0658) (± 0.0660)	−0.4196 (± 0.2953) (± 0.3107)	−0.4663 (± 0.2885) (± 0.2992)	−0.2965 (± 0.1842) (± 0.3549)
ASSE (dB)		−22.84	−11.35	−10.71	−8.27

FIGURE 4: Effect of noise level on the ASSE for an impulse-train excited system.

5.2. An Application for Vocal Tract System Identification. As a practical application of the proposed method, the identification of a vocal tract system is performed from natural speech signals. Since, in this case, the true system parameters are not known, for the purpose of evaluating the estimation accuracy, nonparametric PSD is used. In addition, an estimate of the poles under a noise-free condition is also obtained by using some commonly used technique for the LPC analysis, such as the MLSYW method. The corresponding wide-band spectrogram of the noise-free speech gives information on possible pole locations. In order to estimate the vocal tract system parameters, some English natural voiced phonemes from the TIMIT and the North Texas standard databases [21, 22] with a sampling frequency of 16 KHz are used as the noise-free output observations. Instances of the phonemes

for the TIMIT database are extracted from the database according to the given transcriptions, and the North Texas is a database containing natural vowels. Low-pass filtering up to a certain high-frequency range, such as 6 KHz, is not performed in order to observe the accuracy of the pole estimation over the entire range of frequency. With the estimated parameters of the vocal tract considered as an AR system and the pitch-period (or the excitation signal), a speech phoneme can be synthesized using an appropriate value of the vocal tract filter gain, which is determined based on the RMS power level and the peak PSD of the natural speech frames [1]. For computing synthesized speech signals by different methods, the same excitation signal is used for a particular phoneme. In order to verify the estimation accuracy, first, the PSD of the synthesized speech is compared with that of the noise-free natural speech, and then the estimated poles at a noisy condition are compared with that obtained in a noise-free condition by using the MLSYW method. Figure 5(a) shows a comparison of the PSDs of the vocal tract system obtained from the different methods considered in noisy environments with respect to noise-free PSD. Considering the fact that the choice of the order of the vocal tract filter depends on the spectral characteristics of the specific phoneme, an AR(10) model is used for a naturally spoken sound /ɛ/ of the word "*head*" uttered by a female speaker. In this case, the vowel duration is 128 ms with 2048 samples. In order to test the performance of the methods in estimating the AR parameters of a vocal tract system, twenty independent experiments were performed by adding to the same original speech 20 different realizations of white Gaussian noise, thus obtaining 20 realizations of noisy observations each with a SNR value of −5 dB. These 20 realizations of the noisy observations are then used one by one in twenty independent experiments. In each experiment, one set of AR parameters is obtained by employing a given method of parameter estimation. The AR parameter values of the vocal tract are averaged over 20 sets and then used to obtain the synthesized speech corresponding to the given
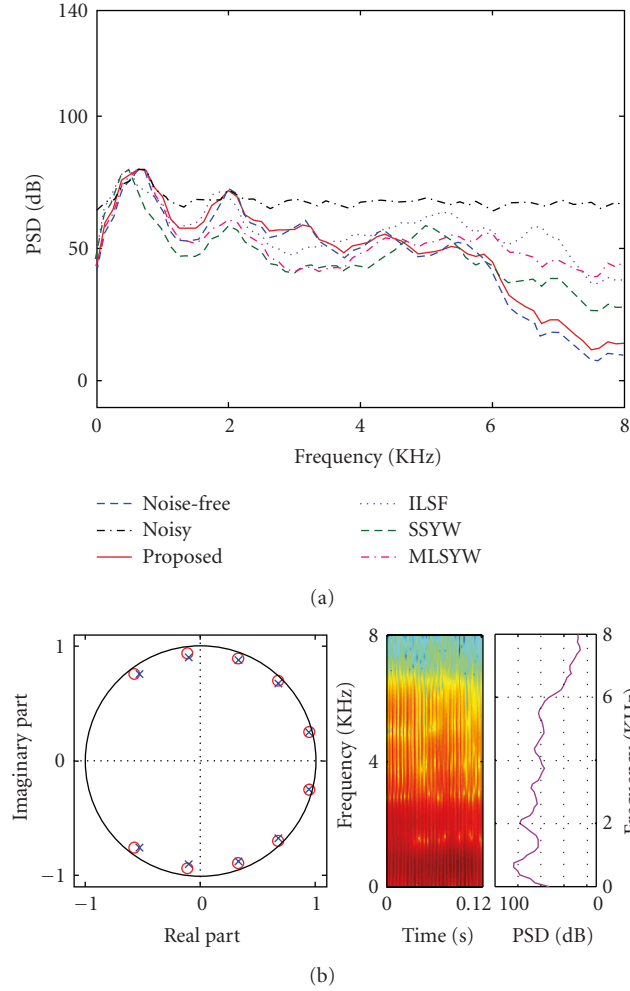
FIGURE 5: Estimation results for a natural speech phoneme /ɛ/ in the presence of white noise at SNR = −5 dB. (a) PSD obtained by using different methods, (b) Average estimated poles (×) obtained from noise-corrupted speech by using the proposed method along with the noise-free estimates (○) obtained by the MLSYW method, spectrogram of the noise-free speech, and noise-free PSD.

FIGURE 6: Estimation results for a natural speech phoneme /a/ in the presence of a multitalker babble noise at SNR = −5 dB. (a) PSD obtained by using different methods, (b) Average estimated poles (×) obtained from noise-corrupted speech by using the proposed method along with the noise-free estimates (○) obtained by the MLSYW method, wide-band spectrogram of the noise-free speech, and noise-free PSD.

method of parameter estimation. We choose the number of RCC samples N_c less than the pitch period T; thus, $N_c = \min(T-1, 10M)$. According to the general behavior of the vocal tract parameter, r_l is searched in the range $[0.8, 0.99]$ [23]. The search range for ω_l can be narrowed down based on the knowledge of the pole locations of a particular phoneme [1, 23]. In order to have a better understanding of the level of noise, the PSD of one of the 20 noisy signals is also included in obtaining the results of Figure 5(a). It is seen from this figure that the PSD of the synthesized signal obtained by using the estimated vocal tact system parameters resulting from the proposed scheme is quite accurate relative to that obtained by the other methods. The estimated average poles are also shown in Figure 5(b) along with the noise-free estimates obtained by the MLSYW method. In Figure 5(b), the noise-free wide-band spectrogram and the noise-free nonparametric PSD are included in order to clearly visualize
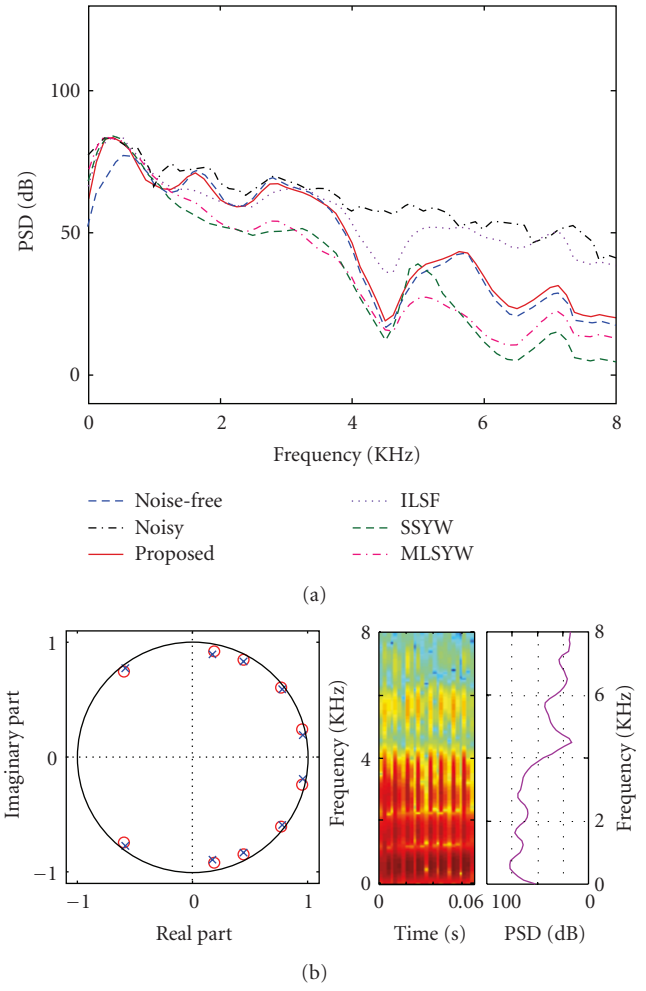
the pole locations and strength in the natural phoneme. The pole-plot clearly shows a high estimation accuracy of the proposed method even at a low level of SNR.

In a similar fashion, using an AR(10) model, PSD results are obtained by employing different schemes under a real noisy environment of a multitalker babble noise (multiple background competing speakers) taken from the Noisex92 database [24]. In Figure 6(a), the results obtained at an SNR of −5 dB for a naturally spoken sound /a/ of the word "*Rob*" uttered by a male speaker are presented. In this case, the vowel duration is 64 ms with 1024 samples. The multiplicity of speakers produces a flatter short-term spectrum which has greater spectral and temporal modulation than a white Gaussian noise. It is observed from Figure 6(a) that the PSD obtained using the proposed method closely matches the noise-free PSD, and all pole locations are accurately estimated. The pole estimation accuracy of the proposed

method is better revealed in Figure 6(b). In this figure, the estimated average poles along with the noise-free pole estimates, the wide-band spectrogram, and the nonparametric PSD are shown. Figure 6 clearly shows that the proposed method is capable of providing a satisfactory estimation performance also in the presence of babble noise at a very low level of SNR.

6. Conclusion

In this paper, a new technique for the parameter estimation of an AR system, given its noise-corrupted output observations, has been proposed. A comprehensive and accurate ramp cosine cepstrum (RCC) model of the one-sided ACF of an AR signal, valid for both white noise and periodic impulse-train excitations, has been developed in a unified fashion in order to identify the AR systems. A residue-based least-squares ramp cosine cepstral fitting scheme employing the RCC model has been presented. It has been shown that the proposed method is able to provide a more accurate estimate of the AR parameters. It combines the attractive features of the correlation- and cepstral-domain system identifications, and has the advantage of providing the flexibility in incorporating some *a priori* knowledge of the parameters, if available, to facilitate the process of parameter estimation. Extensive experimentation performed on different AR systems has demonstrated that the proposed method is sufficiently accurate and consistent in estimating the parameters of the AR signals at very low levels of SNR. The method has also been applied to noise-corrupted natural speech signals for the estimation of human vocal tract system parameters, the accuracy of which is demonstrated in terms of the PSD of the resulting synthesized speech. The simulation results have revealed that the proposed method is superior to some of the existing methods in handling the parameter estimation problem of natural speech signals under white or real-life babble noise degradation.

Acknowledgments

References

[1] D. O'Shaughnessy, *Speech Communications Human and Machine*, IEEE Press, New York, NY, USA, 2nd edition, 2000.

[2] S. M. Kay, *Modern Spectral Estimation, Theory and Application*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[3] N. Xie and H. Leung, "Blind identification of autoregressive system using chaos," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 9, pp. 1953–1964, 2005.

[4] L. Vergara-Dominguez, "New insights into the high-order Yule-Walker equations," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 9, pp. 1649–1651, 1990.

[5] C. E. Davila, "A subspace approach to estimation of autoregressive parameters from noisy measurements," *IEEE Transactions on Signal Processing*, vol. 46, no. 2, pp. 531–534, 1998.

[6] W. X. Zheng, "Fast identification of autoregressive signals from noisy observations," *IEEE Transactions on Circuits and Systems II*, vol. 52, no. 1, pp. 43–48, 2005.

[7] Z. Huang, X. Yang, X. Zhu, and A. Kuh, "Homomorphic linear predictive coding. A new estimation algorithm for all-pole speech modelling," *IEE Proceedings, Part I: Communications, Speech and Vision*, vol. 137, no. 2, pp. 103–108, 1990.

[8] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Identification of autoregressive systems in noise based on a ramp-cepstrum model," *IEEE Transactions on Circuits and Systems II*, vol. 55, no. 10, pp. 1051–1055, 2008.

[9] F. Wang and P. Yip, "Cepstrum analysis using discrete trigonometric transforms," *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 538–541, 1991.

[10] M. Athineos and D. P. W. Ellis, "Autoregressive modeling of temporal envelopes," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5237–5245, 2007.

[11] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[12] S. J. M. de Almeida, J. C. M. Bermudez, N. J. Bershad, and M. H. Costa, "A statistical analysis of the affine projection algorithm for unity step size and autoregressive inputs," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 7, pp. 1394–1405, 2005.

[13] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "A novel technique for the identification of ARMA systems under very low levels of SNR," *IEEE Transactions on Circuits and Systems I*, vol. 55, no. 7, pp. 1988–2001, 2008.

[14] D. G. Long, "Exact computation of the unwrapped phase of a finite-length time series," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 11, pp. 1787–1790, 1988.

[15] W. Verhelst and O. Steenhaut, "A new model for the short-time complex cepstrum of voiced speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 1, pp. 43–51, 1986.

[16] T. Kobayashi and S. Imai, "Spectrum analysis using generalized cepstrum," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 5, pp. 1087–1089, 1984.

[17] T.-H. Hwang, L.-M. Lee, and H.-C. Wang, "Cepstral behaviour due to additive noise and a compensation scheme for noisy speech recognition," *IEE Proceedings Vision, Image and Signal Processing*, vol. 145, no. 5, pp. 316–321, 1998.

[18] H. K. Kim and R. C. Rose, "Cepstrum-domain acoustic feature compensation based on decomposition of speech and noise for ASR in noisy environments," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 5, pp. 435–446, 2003.

[19] C. I. Byrnes, P. Enqvist, and A. Lindquist, "Cepstral coefficients, covariance lags, and pole-zero models for finite data strings," *IEEE Transactions on Signal Processing*, vol. 49, no. 4, pp. 677–693, 2001.

[20] J. Hernando and C. Nadeu, "Linear prediction of the one-sided autocorrelation sequence for noisy speech recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 80–84, 1997.

[21] J. S. Garofolo, L. F. Lamel, W. M. Fisher, et al., "Timit acoustic-phonetic continuous speech corpus," in *Proceedings of Linguistic Data Consortium*, Philadelphia, PA, USA, 1993.

[22] J. M. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *Journal*

of the Acoustical Society of America, vol. 97, no. 5 I, pp. 3099–3111, 1995.

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

[24] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Communication*, vol. 12, no. 3, pp. 247–251, 1993.