AR AND ARMA SYSTEM IDENTIFICATION TECHNIQUES UNDER HEAVY NOISY CONDITIONS AND THEIR APPLICATIONS TO SPEECH ANALYSIS

Shaikh Anowarul Fattah

A thesis
in
The Department
of
Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy
Concordia University
Montréal, Québec, Canada

September 2008

Your file Votre référence
ISBN: 978-0-494-45658-3
Our file Notre référence
ISBN: 978-0-494-45658-3

Abstract

AR and ARMA System Identification Techniques under Heavy Noisy Conditions and their Applications to Speech Analysis

Shaikh Anowarul Fattah, Ph.D.

Concordia University, 2008

System identification under noisy environment has axiomatic importance in numerous fields, such as communication, control, and signal processing. The system identification is to estimate and validate the parameters of the system from its output observations, a task that becomes very difficult when the system output is heavily noise-corrupted. The major objective of this research is to develop novel system identification techniques for an accurate estimation of the parameters of minimum phase autoregressive (AR) and autoregressive moving average (ARMA) systems in practical situations where the system input is not accessible and only noise-corrupted observations are available. Unlike conventional system identification methods in which only the white noise excitation is considered, both the white noise and periodic impulse-train excitations are taken into account in the methodologies developed with an aim of directly using them in speech analysis.

A new ARMA correlation model is developed, based on which a two-stage correlation-domain ARMA system identification method is proposed. In the first stage, the new model in conjunction with a residue based least-squares (RBLS) model-fitting optimization algorithm is used to estimate the AR parameters. In the second stage, the moving average (MA) parameters are estimated from the residual signal obtained by filtering the observed data using the estimated AR parameters. With a view to overcome the adverse affect of noise on the MA part, a noise-compensation scheme using

an inverse autocorrelation function (IACF) of the residual signal is also proposed.

Cepstrum analysis has been popular in speech and biomedical signal processing. In this thesis, several cepstral domain techniques are developed to identify AR and ARMA systems in noisy conditions. First, a ramp-cepstrum model for the one-sided autocorrelation function (ACF) of the AR and ARMA signals is proposed, which is then used for the estimation of the parameters of AR or ARMA systems using the RBLS algorithm. It is shown that for the estimation of the MA parameters of the ARMA systems, either a direct ramp-cepstrum model-fitting based approach or a noise-compensation based approach can be adopted. Considering that, in the case of real signals, discrete cosine transform is more attractive than the Fourier transform (FT) in terms of the computational complexity, a ramp cosine cepstrum model is also proposed for the identification of the AR and ARMA systems.

In order to overcome the limitations of the conventional low-order Yule-Walker methods, a noise-compensated quadratic eigenvalue method utilizing the low-order lags of the ACF, is proposed for the estimation of the AR parameters of the ARMA system along with the noise variance. For the estimation of the MA parameters, the new noise-compensation method,in which, a spectral factorization of the resulting noise-compensated ACF of the residual signal is used, is employed.

In order to study the effectiveness of the proposed identification techniques, extensive simulations are carried out by considering synthetic AR and ARMA systems of various orders under heavy noisy conditions. The results demonstrate the significant superiority of the proposed techniques over some of the existing methods even under very low levels of SNR. Simulation results on the identification of human vocal-tract systems using natural speech signals are also provided, showing a superior performance of the new techniques.

As an illustration of application of the proposed AR and ARMA system identification techniques to speech analysis, noise robust schemes for the estimation of formant frequencies are developed. Synthetic and natural phonemes including some naturally spoken sentences in noisy environments are tested using the new formant estimation schemes. The experimental results demonstrate a performance superior to that of some of state-of-the-art methods at low levels of SNR .

Dedication

To my beloved wife Celia Shahnaz

Acknowledgments

I would like to express my sincere gratitude and profound indebtedness to my supervisors Dr. M. Omair Ahmad and Dr. Weiping Zhu for their guidance, encouragement, constructive suggestions, and support during the span of this research. I am grateful to them for providing me the freedom and motivation to explore new areas of research and new ideas. I also want to thank them for spending so many hours with me in correcting and improving the writing of this thesis. The useful suggestions provided by the committee members are also deeply appreciated.

I want to acknowledge the financial support provided by Concordia University and NSERC, Canada, which was crucial for completing this research.

I wish to thank my colleagues in the Center for Signal Processing and Communications Dr. I. H. Bhuiyan, Mr. Awni Itradat, Dr. Chao Wu, and Dr. Saad Boguezel for their friendship and moral support.

Special note of thanks goes to my wife, Ms. Celia Shahnaz, for her continuous moral support, inspiration, thoughtful comments, and friendly cooperation.

Finally, I would like to express my gratefulness to my parents, and in-laws for their patience, love and constant encouragement.

Table of Contents

List of Figures

List of Tables

List of Acronyms

ACF	:	Autocorrelation function
ACR	:	ARMA-cepstrum recursion
ACM	:	ARMA correlation matching
AFB	:	Adaptive filter bank
AR	:	Autoregressive
ARMA	:	Autoregressive moving average
ARMAC	:	ARMA correlation
ARMARC	:	ARMA ramp-cepstrum
ARRC	:	AR ramp-cepstrum
ASSE	:	Average sum-squared error
AWGN	:	Additive white Gaussian noise
CC	:	Complex cepstrum
DCT	:	Discrete cosine transform
DFT	:	Discrete Fourier transform
DYW	:	Durbin's MA with LSMYW for AR parameter estimation
FT	:	Fourier transform
FFT	:	Fast Fourier transform

HOYW	:	Higher-Order Yule-Walker
IACF	:	Inverse Autocorrelation Function
IAR	:	Intermediate AR
IDCT	:	Inverse discrete Cosine transform
IFFT	:	Inverse Fast Fourier transform
ILSD	:	Improved least-squares with direct implementation structure
ILSF	:	Improved least-squares with faster convergence
LOYW	:	Lower-Order Yule-Walker
LPC	:	Linear predictive coding
LS	:	Least-squares
LSMYW	:	Least-squares MYW
LSYW	:	Least-squares YW
LTI	:	Linear time-invariant
MA	:	Moving average
ML	:	Maximum likelihood
MYW	:	Modified Yule-Walker
NCCD	:	Noise compensated correlation domain
ORACF	:	Once-repeated autocorrelation function
OS	:	One-step method

OSACF	:	One-sided autocorrelation function
OSD	:	Order-selective Durbin
PSD	:	Power spectral density
RBLS	:	Residue based least-squares
RC	:	Ramp cepstrum
RCC	:	Ramp cosine cepstrum
RMSE	:	Root mean square error
RS	:	Reduced statistics
SAPE	:	Spectral all-pole estimation
SDM	:	Standard deviations from mean
SDT	:	Standard deviations from true values
SI	:	System identification
SNR	:	Signal-to-noise ratio
SSYW	:	Signal/noise subspace based Yule-Walker
TIMIT	:	Texas instruments/Massachusetts institute of technology
TS	:	Two-stage method
VT	:	Vocal tract
YW	:	Yule-Walker

Chapter 1

Introduction

1.1 General

Estimation of system parameters from given observations is concerned with characterizing the system behavior. Its main tasks are the estimation and validation of parametric models for the systems. It has axiomatic importance in various fields, such as signal processing, communication, and control [1]. For example, blind channel identification, without knowing the input data, offers a way of reducing intersymbol interferences in digital communication [2], [3]. Control engineers often use model-based identification techniques to gain physical insights into the underlying dynamics of the systems [4]–[6]. In the area of signal processing, system identification has widespread applications, covering speech processing, biomedical signal processing, image processing, radar and seismic signal processing. For instance, estimation of speech production system parameters plays an important role in digital speech analysis/synthesis, modeling, coding, and recognition [7]–[12]. In biomedical signal processing, it is a common practice to fit physiological systems with some simplified models for the diagnosis of some physical phenomena [13]–[16]. In the area of radar signal processing, a model-based spectral analysis is very frequently used [17]–[19],

and in the seismic signal processing, system identification methods are employed to estimate different parameters, such as the reflection coefficients [20].

A parametric model extracts the important aspects of the dynamic behavior of the system under test. In many practical applications, discrete-time systems are well approximated by rational transfer function models, which, in general, can be classified into three categories: (1) the pole-zero or autoregressive moving average (ARMA) model containing both poles and zeros, (2) the all-pole or autoregressive (AR) model containing only poles, and (3) the all-zero or moving average (MA) model containing only zeros. In comparison to the other two models, the ARMA model provides a more efficient way of representing the system in terms of the required number of model parameters and spectral matching characteristics. The AR model is in general suitable for representing spectra with narrow peaks, which commonly appear in the modeling of real-life signals,while if using the MA model, a large number of coefficients are required to represent the narrow spectrum. System parameters are usually estimated from the given output observations. Fig. 1.1 shows a typical system identification problem in the presence of noise, where the data available for estimating the system parameters are the noise-corrupted output observations. This figure helps drawing our attention to some major issues involved in the parameter estimation problems that are summarized as follows:

1. In most of the real-life applications, the input data to the system are not available or not accessible but some of its properties could be conjectured. Obviously, if the input signal can be accessed, the parameter estimation becomes a more standard and simpler problem.

2. System transfer function can often be modeled in advance as an ARMA, AR, or

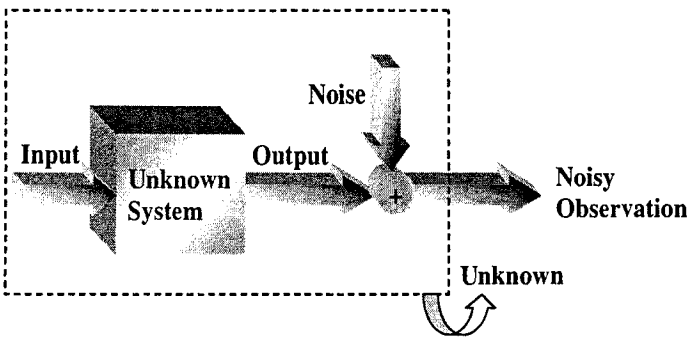


Figure 1.1: System identification in the presence of noise.

MA models depending on the practical applications, but the model parameters are unknown in a particular identification test.

3. In the parameter estimation problem, in most of the cases, only the noise-corrupted output observations are available. Although the noise-only part of the output is generally not available, some of its statistical properties can be conjectured.

4. The level of noise corrupting the system output, which is measured in terms of the signal-to-noise ratio (SNR), significantly affects the estimation performance.

In the real world, it is quite impossible to avoid the inclusion of environmental noise. As far as real-life applications are concerned, parameter estimation using only the noise-corrupted observations at a very low SNR is rather difficult. For example, in the case of speech signals, the identification problem becomes very tough, especially at a very low SNR level, since the input data are not available or accessible and the estimation task has to be performed only from a given noisy observation. Thus, the parameter estimation from a very severe noisy observation is an open problem that has interested many researchers.

1.2 A Review of Existing Methods of System Identification

In this section, we will review some of the typical techniques for the estimation of system parameters. This literature review not only serves as a necessary background material in understanding the state-of-the-art techniques, but also supports the motivation of the research work in the thesis.

1.2.1 AR and ARMA System Identification

In most of the application fields, as mentioned in Section 1.1, the parameter estimation problems mainly consider either the ARMA model [2], [5], [11]–[14], [21]–[23] or the AR model [3], [4], [6], [9], [10], [16]–[18], [24]. Among several methods of parameter estimation, the maximum likelihood (ML) or pseudo-ML algorithms were extensively studied for the identification of AR [25]–[30] and ARMA [31]–[34] systems. Most of the ML estimation methods only deal with the noise-free condition. The ML methods are asymptotically consistent but their convergence performance relies heavily on the initialization process of the methods [33]–[35]. Recently in [35], a genetic algorithm has been proposed to solve the ML estimation problem in the presence of noise. In this method, unlike conventional estimation problems, the AR system driven by chaos has been considered, which results in a better estimation accuracy compared to that obtained by using the white Gaussian probing signal with the least-squares estimator [36], [37]. Note that the estimation based on a probing signal requires an access to the system-input which, as mentioned in the previous section, is not possible for most of the real-life applications.

Instead of simultaneous estimation of the AR and MA parameters of the ARMA systems, most of the methods estimate the parameters separately. These two-stage

methods offer a better controllability on the estimation accuracy. Estimation of the AR parameters of the ARMA systems, if performed at the first stage, is of crucial importance for the MA parameter estimation at the second stage, since a poor AR estimation result would severely affect the estimation accuracy of the MA part. For the first stage of ARMA system identification, the autocorrelation function (ACF) based algorithms have been most widely used such as the modified/extended Yule-Walker (MYW) methods [33], [38]–[47]. The main advantage of this approach is that it requires only a simple set of linear equations involving sample autocorrelation coefficients. In a noisy environment, the estimation performance of the MYW methods gets significantly degraded due to the presence of errors in all the lags of the ACF of the noisy signal.

It is to be mentioned that, different variants of Yule-Walker (YW) algorithms have also been extensively employed to identify the AR systems [33], [48]. The estimation performance of noise compensation based identification schemes, such as the low-order YW (LOYW) method, strongly depends on the accuracy of the *a priori* knowledge of the noise corrupting the signals [33], [49]–[51]. Although the high-order YW (HOYW) method does not require *a priori* estimate of the noise variance, it suffers from a singularity problem and has a large estimation variance [52]–[54]. In this case, to reduce the estimation variance, a least-squares HOYW (LSYW) method can be used [33], [55]. However, in the presence of a moderate level of noise, the estimation variance of the LSYW method is still large [56], [57]. In order to overcome this problem, a signal/noise subspace YW (SSYW) method has been proposed by introducing a noise compensation in the LOYW method [58], [59]. The least-squares (LS) based techniques for the AR parameter estimation have also been frequently used [60]–[65]. They involve less computation but performs well only at a high level

of SNR. The methods reported in [66], [67], have claimed a better identification of AR systems at a low level of SNR. However, the method in [66] involves nonlinear operation and suffers from a convergence problem. In [67], a pre-filtering of the estimated energy density spectrum in the discrete cosine transform (DCT) domain is utilized to estimate the noise-free autocorrelation function which is used for the AR parameter estimation. This method strongly suffers from the problem of selecting the unknown noise level to be subtracted. In addition to this, some assumptions like linearity of the square DCT coefficients and the uncorrelation between signal and noise DCT coefficients introduce significant error in the parameter estimation.

Using the estimated AR parameters, one can filter the ARMA signal and obtain a residual signal that can be utilized to estimate the MA parameters of the ARMA systems with the help of some suitably chosen MA estimators [33], [68]. There exist mainly two approaches to estimate the parameters from the MA residual signal: (1) the spectral factorization [44] and (2) the Durbin's algorithm [33]. Both of these approaches severely suffer under noisy conditions, since the observation noise heavily corrupts the residual signal. The estimation performance of the Durbin's method of MA parameter estimation strongly depends on both the accuracy of the estimated intermediate AR (IAR) parameters and the order chosen for the IAR system [33]. Instead of using the IAR model, in [43] and [45], some direct schemes have been proposed to estimate the MA parameters using the residual signal. However, the method described in [43] needs to solve a set of nonlinear equations and the method in [45] has been tested only in the noise-free condition. Note that the noise-part corrupting the residual signal is also an MA signal, making the estimation problem very difficult, especially when the SNR is very low. In order to alleviate the noise effect on the MA parameter estimation of the residual signal, one possible solution,as

will be seen from this thesis work, is to use an effective noise reduction technique operating on the noise-corrupted residual signal prior to the MA estimation.

In speech and biomedical signal processing, cepstrum analysis has been proved to be extremely effective [69]–[72]. Estimating system parameters from the cepstral coefficients has been attempted by a few researchers [8], [11], [45], [73], [74]. Among them, methods described in [11], [45], [74] are proposed for the identification of the noise-free ARMA systems. In [45], an ARMA-cepstrum recursion (ACR) is proposed using the estimated cepstral coefficients of the ARMA signal in order to estimate the MA parameters. The method proposed in [11] involves spectral factorization and minimization of a convex objective function. Recently in [74], an enhanced cepstrum is first computed by using the eigen-decomposition of the covariance matrix of the observed data and then utilized in a conventional method of ARMA parameter estimation. The cepstrum based noise-free AR estimation technique, proposed in [73], cannot guarantee the stability of the estimated AR model.

Identification of ARMA systems under noisy conditions, especially using only the output observations, has been attempted by some researchers [75]–[77]. Among them, the method presented in [75] is a lattice filter based method, which provides good estimates for the AR part but fails to accurately identify the MA part in the presence of noise. Methods proposed in [76], [77] estimate the AR parameters from the ACF of noisy observations and they estimate the MA parameters from the ACF using an autocorrelation matching (ACM) technique under the assumption that the excitation power is known. Furthermore, the estimation accuracy of the ACM method [77] is affected in the presence of heavy noise due to its requirement of using high order models equivalent to the noisy ARMA process as well as due to the employment of the lattice filter based method.

There are some recently developed methods that deal with noisy environments where the input signal needs to be accessed or designed [13], [14], [78], [79]. In order to overcome the problem associated with the noisy observed signal, in [78], by designing input as a chaotic signal, the ergodic property of the dynamical system has been exploited, while in [80], a sinusoidal probing signal has been used with an appropriate amplitude estimator. In [79], a technique combining the time and frequency domain approaches has been used to estimate the ARMA parameters. Methods in [14] and [13] identify the ARMA systems by minimizing a suitably chosen cost function. In [14], a better estimation accuracy has been achieved in the noise-free case but the performance is degraded in the presence of noise. The method proposed in [13], gives a biased estimation in situations where the number of observed samples or the signal-to-noise ratio (SNR) is low.

1.2.2 Application of the Parameter Estimation Methods in Speech Processing

In recent years, there has been an increasing demand for the development of accurate, efficient, and compact representations of speech production systems. Such representations require the extraction of the characteristics of a vocal-tract system from speech signals. Thus, vocal-tract SI has received potential applications in many areas of speech processing, such as speech analysis/synthesis, speech coding, speech recognition, acoustics phonetics, modeling of speech production process [81]. Formants are defined as free resonances of a human vocal-tract system [7]. They serve as an important acoustic feature and offer phonetic reduction in speech recognition [82]. It also plays an important role in the design of some hearing aids [83]. Furthermore, formant-based speech synthesis analysis is nowadays receiving more and more

research attentions because of its continuity in unit concatenation.

Among different formant estimation techniques, linear predictive coding (LPC) based methods are most commonly used, where formant frequencies are determined through the LPC spectral analysis that is normally done by peak picking or pole finding [7], [84], [85]. However, these methods characteristically offer a moderate noise immunity only.Cepstrum has also been used in the formant estimation [86], [87]. In [87], vocal-tract resonances are estimated using the LPC cepstral coefficients. In this method, the resonance candidates are obtained by employing a full-space search of all frequency values in order to fit the measured acoustic data to both the nonlinear prediction function and the temporal constraint simultaneously. It has been demonstrated that the LPC cepstrum is advantageous over the LPC spectrum because it provides a much simpler form of the analytical function. Like most of the formant estimation methods so far reported, this method has dealt only with noise-free environments [84], [85], [87]–[92].

In real-life applications, formant estimation from noise-corrupted speech is essential but a difficult task. As a matter of fact, only a few research results concerning formant estimation from noisy speech are available in literature [93], [94]. The method in [93] provides a consistent formant frequency estimates at a relatively high SNR by a pitch-synchronous averaging of the covariance estimates over a number of consecutive pitch periods. In [94], a peak-picking algorithm was used to a segmented spectrum to estimate the formant frequencies. This method offers advantage of determining the segment boundaries sequentially but the segmentation algorithm requires the ACF of the clean signal which restricted the performance up to an SNR of 5 dB. Recently, based on the filtering principle proposed in [95], an adaptive filter-bank (AFB) method was proposed in [83], which estimates a formant frequency from the

corresponding band-pass filtered output observations by using the conventional autocorrelation method. This method provides an incorrect formant estimation if the initial estimates are far from the actual values or if the SNR is very low.

1.3 Motivation

From the above discussion, the estimation performance of most of the AR and ARMA system identification methods, whether they have been developed in the correlation domain or in the cepstral domain, deteriorates drastically in the presence of noise. The situation becomes worse when the estimation task of an ARMA system identification has to be performed according to the noise-corrupted output observations. As for practical applications, there has been a growing demand for the noise-robust system identification schemes, especially in the area of speech signal processing. Thus, there is a need of making concerted effort towards the developments of the identification algorithms that can provide accurate parameter estimation of AR and ARMA systems in practical situations where there exists noise or even the SNR is very low.

Although some methods have already been proposed to handle the ARMA system identification problems in noisy environments, they have a common shortcoming of requiring an access to the input excitation. It is to be mentioned that in many real-life applications involving speech or biomedical signals, the input excitation is not available or accessible. Therefore, the applications of the existing system identification methods are very much limited.

Since most of the system identification methods in literature cannot effectively deal with natural signals, new methods must be developed that by using appropriate input signals can create a situation in which the natural signals could be handled more efficiently. In most of the conventional system identification methods, only the white

noise excitation is considered. However, in order to handle the problem of human vocal-tract system identification or speech signal analysis, methods must be devised by considering both white noise and periodic impulse-train excitations in order to create a natural-signal-like situation.

One of the criteria used in system identification problems is to minimize, in the least-square sense, the error between the true data and the noise-corrupted observed data. Since true data is not available in practice, an appropriate model, if it exists, to represent the true data will facilitate the least-squares estimation problem. This idea could be used for developing noise-robust models in the correlation and cepstral domains, which could then be used to develop the least-squares model-fitting based identification techniques. In view of the fact that generally white noise or periodic impulse-train is used as input to naturally excited system, such as human vocal-tract system, any new system identification must be developed based on both the excitations. Finally, the performance evaluation and validation of any new system identification technique must be supported by its real-life applications.

1.4 Scope and Organization of the Thesis

The main objective of this research is to develop effective methodologies for the identification of AR and ARMA systems, which are able to estimate the system parameters accurately using the noise-corrupted observations even at a very low level of SNR. Unlike some state-of-the-art techniques, the input to the system is not assumed to be accessible or available. In order to show the effectiveness of the proposed techniques in handling real-life applications, especially in the area of speech signal processing, both white noise and periodic impulse-train excitations are considered in the system identification techniques. We intend to develop effective noise-robust methods in the

correlation- and cepstrum-domains by utilizing their advantageous features. In order to demonstrate the effectiveness of the proposed methods, extensive simulations are performed by considering synthetic AR and ARMA systems of different orders under noisy conditions and the results are compared with those from some of the existing methods. Simulation results are also provided for the identification of human vocal-tract systems using natural speech signals. As a practical application of the system identification techniques proposed in this thesis, experimental results on the formant frequency estimation of different synthetic and natural speech signals are presented. The thesis is organized as follows.

In Chapter 2, a new correlation model for the ARMA signal is proposed, based on which an ARMA system identification method is developed. With a view to achieve the suitability of the proposed model in speech processing applications, in the derivation of the ARMA correlation (ARMAC) model, both the periodic impulse-train and the white Gaussian noise inputs are taken into consideration. The ARMAC model is then employed in developing a two-stage ARMA system identification scheme. In the first stage, a residue-based least-squares (RBLS) correlation fitting optimization algorithm is presented for the estimation of the AR parameters. Identification of the MA part of the ARMA system under a heavy noisy condition is a very difficult task. With a view to solve this problem, in the second stage, the noise in the residual signal is first reduced by using a new noise-reduction scheme and then, the problem of the MA parameter estimation is converted into a correlation fitting scheme by using the inverse ACF (IACF) corresponding to the resulting noise-compensated residual signal.

In Chapter 3, a cepstral-domain system identification method is proposed based on a new ramp-cepstrum model for the one-sided autocorrelation function (OSACF)

of the observed signal. Considering that very few cepstral-domain system identification methods are available in literature, we first present the proposed AR system identification technique and then extend it for ARMA systems. With regard to the AR system identification, a ramp-cepstrum (RC) model for the OSACF of the AR signal is first proposed by considering both types of inputs. As an extension of this new idea, an ARMA ramp-cepstrum model is also derived. A ramp-cepstral model fitting approach with the RBLS algorithm developed in Chapter 1 is then proposed to estimate the AR parameters of the AR or ARMA system from the noisy output observations. For the identification of the MA part of the ARMA system, we propose two methods. The first method directly estimates the MA parameters using the RC model in a similar manner as the AR parameter are estimated. In the second method, the MA parameters are estimated from the residual signal by using a noise-reduction scheme.

In Chapter 4, another cepstral-domain system identification method is developed based on cosine cepstrum, where a new ramp-cosine cepstrum (RCC) model for the OSACF of the observed signal, considering both types of inputs, is proposed. At first, we present the proposed RCC based method for the AR system identification and then extend it for the ARMA system. In order to estimate the AR parameters, the RBLS algorithm, which guarantees the stability of the system, is employed in conjunction with the RCC model. For the purpose of implementation, the discrete cosine transform (DCT), which can efficiently handle the phase unwrapping problem and offer computational advantages over the discrete Fourier transform, is employed.

Apart from the model-fitting based identification techniques developed in the previous chapters, in Chapter 5, we propose a new noise-compensated quadratic eigenvalue method for the estimation of the AR parameters of the ARMA system in order

to overcome the limitations of the conventional low-order Yule-Walker methods. At the first stage, the AR parameters along with the noise variance are estimated by solving a quadratic eigenvalue problem corresponding to a set of equations containing the lower lags of the ACF of the observed data. The MA parameters are then estimated at the second stage by using the spectral factorization of a noise-compensated ACF of the residual signal.

Chapter 6 is concerned with formant estimation based on the proposed system identification techniques. First, we present a general framework for formant frequency estimation which consists of some functional blocks such as windowing, preprocessing, pre-filtering etc. We then propose some robust formant estimation methods which can efficiently tackle the adverse effect of observation noise, including (1) ARMA correlation method, (2) ramp-cepstrum (RC)method, and (3) ramp cosine cepstrum (RCC)method. In the ARMA correlation method we use the ARMAC model in conjunction with an adaptive RBLS algorithm to extract first few formant frequencies from the observed noise-corrupted speech. The RC method estimates the formant frequencies from noisy speech by employing the proposed RC model in the adaptive RBLS algorithm. The RCC method is developed based on the ramp cosine cepstrum of the once repeated ACF (ORACF). In comparison with the OSACF, the ORACF offers more noise robustness which enhances the estimation performance of the RCC based formant estimation method.

Finally, some concluding remarks highlighting the contributions of the thesis and suggestions for future work are provided in Chapter 7.

Chapter 2

ARMA Correlation Model Based System Identification

2.1 Introduction

Identification of ARMA systems under a heavy noisy condition, especially when using only the output observations, is a very difficult but essential task for many practical applications. In a noisy environment, the performance of the correlation based methods gets significantly degraded due to the presence of errors in all the lags of the autocorrelation function (ACF) of the noisy signal. In this chapter, a novel technique for the identification of minimum-phase autoregressive moving average (ARMA) systems from the output observations heavily corrupted by noise is presented [96], [97]. First, starting from the conventional correlation estimator, a simple yet accurate ARMA correlation (ARMAC) model in terms of the poles of the ARMA system is presented in a unified manner for white noise and impulse-train excitations. The AR parameters of the ARMA system are then obtained from the noisy observations by developing and using a residue-based least-squares (RBLS) correlation-fitting optimization algorithm that employs the proposed ARMAC model. As for the estimation of the MA parameters, it is preceded by the application of a new technique intended to reduce

the noise present in the residual signal that is obtained by filtering the noisy ARMA signal via the estimated AR parameters. A scheme is then devised whereby the task of MA parameter estimation is transformed into a problem of correlation-fitting of the inverse autocorrelation function (IACF) corresponding to the noise-compensated residual signal. In order to demonstrate the effectiveness of the proposed method, extensive simulations are performed by considering synthetic ARMA systems of different orders in the presence of additive white noise and the results are compared with those of some of the existing methods. It is shown that the proposed method is capable of estimating the ARMA parameters accurately and consistently with guaranteed stability for signal-to-noise ratio (SNR) levels as low as -5 dB. Simulation results for the identification of a human vocal-tract system using natural speech signals are also provided showing a superior performance of the proposed technique in terms of the power spectral density of the synthesized speech signal.

The rest of the chapter is organized as follows. In Section 2.2, we first present the derivation of the ARMAC model for the two types of input excitations. Then, a scheme for the estimation of the AR parameters of the ARMA system in the presence of noise is proposed. Section 2.3 presents a methodology to estimate the MA parameters of the ARMA system. The performance of the proposed method is demonstrated in Section 2.4 through extensive computer simulations for both synthetic and natural speech signals. Finally, in Section 2.5, the salient features of this investigation are summarized.

2.2 Estimation of AR Parameters

2.2.1 Problem Statement

A causal, stable and linear time-invariant (LTI) ARMA (P, Q) system can be characterized by

$$\sum_{i=0}^{P} a_i x(n-i) = \sum_{j=0}^{Q} b_j u(n-j) \tag{2.1}$$

where $u(n)$ and $x(n)$ are, respectively, the excitation and the response of the system, a_i and b_j the corresponding AR and MA parameters with $a_0 = 1$ and without loss of generality we can assume $b_0 = 1$. The orders of the ARMA model P and Q are assumed to be known. If the orders are not known, before starting the task of system identification, an estimate of the system order must be obtained. Different standard techniques are available in the literature [98], [99] that can be employed to estimate the order of an ARMA system. It is also assumed that $P > Q$. Note that, for real-life data, such as speech signal, it is sufficient to use an ARMA model with less number of zeros than the number of poles [7]. Moreover, since the presence of an additive noise directly affects the system zeros, modeling real data with less number of zeros would be helpful in reducing the estimation error. In most of the system identification problems, $u(n)$ is considered to be a stationary, zero-mean white Gaussian noise with variance σ_u^2. For some practical applications, however, the excitation can have other forms [7], [100]. For example, in speech processing, an impulse-train is often used as the excitation of the vocal-tract system [7]. In our identification scheme, both white noise and impulse-train excitations are considered. The transfer function of the

ARMA(P,Q) system described by (2.1) can be written as

$$H(z) = \frac{B(z)}{A(z)} = \frac{\prod_{j=1}^{Q}(1 - z_j z^{-1})}{\prod_{k=1}^{P}(1 - p_k z^{-1})} = \sum_{k=1}^{P} \frac{\eta_k}{1 - p_k z^{-1}} \tag{2.2}$$

where $A(z) = \sum_{k=1}^{P} a_k z^{-k}$ and $B(z) = 1 + \sum_{j=1}^{Q} b_j z^{-j}$ are, respectively, the AR and MA polynomial, p_k's and z_j's denote, respectively, the poles and the zeros of the ARMA system, and η_k the partial fraction coefficient corresponding to the kth pole. As in most of the ARMA system identification methods, it is assumed that all poles and zeros are of the first-order, no further pole-zero cancelation possible, and the ARMA(P,Q) process is minimum phase and stationary. It is to be mentioned that the higher order poles and zeros deal with the cases having sharp peaks or notches in their frequency responses. However, in the presence of noise, modeling with high-order poles and zeros may lead to an ambiguous identification results. In order to avoid this problem and still deal with situations of sharp responses, one can use first-order poles and zeros with appropriate distribution of their locations. Since in many applications dealing with real-life data, only the amplitude response of the system or the corresponding PSD is a major concern, the assumption of minimum phase would not be critical. The impulse response of the ARMA system can be easily obtained from (2.2) as

$$h(n) = \sum_{k=1}^{P} \eta_k p_k^n \tag{2.3}$$

The ACF of $x(n)$ is given by

$$\rho_x(\tau) = E[x(n)x(n+\tau)] \tag{2.4}$$

Given a finite set of observations $\{x(n)\}_{n=0}^{N-1}$, the ACF of $x(n)$ can be estimated as [33]

$$r_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} x(n)x(n+|\tau|), \tau = 0, 1, \ldots, M-1 \tag{2.5}$$

In the correlation-based system identification methods, (2.5), hereafter referred to as the conventional estimator, is generally used to estimate the AR parameters. In practical applications, where data length N is finite, (2.5) offers an efficient way to obtain an accurate estimate of the true ACF $\rho_x(\tau)$ for AR and ARMA signals. In (2.5), a small number, M, of ACF lags is sufficient to represent an accurate estimate of $\rho_x(\tau)$.

In the presence of additive white Gaussian noise $v(n)$, the observed signal $y(n)$ can be written as

$$y(n) = x(n) + v(n) \tag{2.6}$$

where $v(n)$ is a stationary process with zero-mean and variance σ_v^2, and is independent of $u(n)$. The estimated ACF $r_y(\tau)$ of the noisy observation $y(n)$ can be computed using (2.5) as

$$r_y(\tau) = r_x(\tau) + r_v(\tau) + r_c(\tau) \tag{2.7}$$

where $r_c(\tau)$ represents the cross-correlation terms. Assuming that $x(n)$ and $v(n)$ are uncorrelated, one can obtain an estimate of $r_x(\tau)$ from $r_y(\tau)$ using the relation

$$r_x(\tau) = \begin{cases} r_y(\tau) - \sigma_v^2 & \tau = 0 \\ r_y(\tau), & \tau \neq 0 \end{cases} \tag{2.8}$$

It is worth mentioning that in a heavy noisy condition, the cross-correlation between the signal and noise may not be equal to zero, and the deviation from zero may be large in the case of a very low SNR. Thus, the estimation of $r_x(\tau)$ using (2.8) may result in a significant error at all lags. This is the reason as to why the autocorrelation based methods provide poor estimates of the AR parameters in a heavy noisy environment.

In order to alleviate this problem, in this section a new correlation model for the ARMA signal is first derived. This model is then used as the target function for the noisy ACF in a correlation-fitting technique that employs an effective residual least-squares optimization method for obtaining quite an accurate estimate of $r_x(\tau)$ from $r_y(\tau)$ even at a very low SNR. The correlation model for the ARMA signals is presented for white noise as well as periodic impulse train excitations with both finite or infinite durations.

2.2.2 ARMAC Model for a White Noise Excitation

For a white noise excitation $u(n)$, the output of an initially relaxed ARMA system can be obtained using (2.3) as

$$x(n) = u(n) * h(n) = \sum_{k=1}^{P} \sum_{m=0}^{n} \eta_k u(m) p_k^{n-m} \tag{2.9}$$

where * denotes the convolution operation. Substituting (2.9) into (2.5), $r_x(\tau)$ for $\tau \geq 0$ can be expressed as

$$r_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-\tau} \sum_{k=1}^{P} \sum_{j=1}^{P} \eta_k \eta_j \sum_{l=0}^{n} \sum_{m=0}^{n+\tau} \xi \tag{2.10}$$

$$\xi = u(l)u(m)p_k^{\tau+n-m} p_j^{n-l},\ \tau = 0, 1, \ldots, M-1$$

As the ACF is a decaying function of τ, it is sufficient to consider a small value for the ACF lags M irrespective of the duration N of the excitation as being finite or infinite [33]. We have found through an extensive experimentation that in the case of a finite-duration excitation, it is sufficient to choose the M as $N/4$. Since the magnitude of the poles of a stable ARMA system is less than unity, the product terms containing larger powers of the poles in (2.10), can, therefore, be neglected and the

expression for ACF given by (2.10) can be simplified as (see Appendix A for details)

$$r_x(\tau) = \sum_{k=1}^{P} \psi_{W_k} p_k^{\tau}, \ \tau = 0, \ 1, \ \ldots\ldots, \ M-1 \tag{2.11}$$

$$\psi_{W_k} = \chi \sum_{j=1}^{P} \frac{\eta_k \eta_j}{1 - p_k p_j} \tag{2.12}$$

$$\chi = \frac{1}{N} \sum_{l=0}^{N-\tau-1} u(l)^2 \tag{2.13}$$

Note that for $N \rightarrow \infty$, χ can be treated as the variance of $u(n)$. It is seen from the above equation that the ARMA correlation (ARMAC) model as given by (2.11) is expressed explicitly in terms of the poles of the ARMA system for both finite- and infinite-duration excitations.

2.2.3 ARMAC Model for a Periodic Impulse-train Excitation

The periodic impulse-train excitation $\{u_i(n)\}_{n=0}^{N-1}$ with a period T can be expressed as

$$u_i(n) = \sum_{m=0}^{\lambda-1} \delta(n - mT), \quad \lambda = \lceil N/T \rceil \tag{2.14}$$

where $\lceil \zeta \rceil$ represents the smallest integer greater than or equal to ζ, and thus, λ is the total number of impulses in the excitation. Using (2.3) and (2.14), for an initially relaxed ARMA system, the output can be obtained as

$$\begin{aligned} x(n) &= u_i(n) * h(n) \\ &= \sum_{k=1}^{P} \sum_{m=0}^{\kappa-1} \eta_k p_k^{n-mT} = \sum_{k=1}^{P} \eta_k p_k^n G_{kn} \end{aligned} \tag{2.15}$$

$$G_{kn} = \sum_{m=0}^{\kappa-1} p_k^{-mT} = \begin{cases} 1, \ n = 0, 1, \ldots T-1 \\ 1 + p_k^{-T}, \ n = T, \ldots, 2T-1 \\ \vdots \\ 1 + \ldots + p_k^{-(\kappa-1)T}, \\ \quad n = (\kappa-1)T, \ldots, \kappa T - 1 \end{cases} \tag{2.16}$$

where $\kappa = \left\lceil \frac{(n+1)}{T} \right\rceil$. In the periodic impulse-train excited systems, all the necessary information required to identify the system parameters lies within the first T lags of $r_x(\tau)$, and hence, it is sufficient to consider $M = T/2$ ACF lags. In most of the system identification applications, the data length can be chosen as $N \geq T/2$. The objective here is to develop a correlation model as it was done for the case of white noise excitation.

Due to the complicated form of the signal model (2.15), which, depending on the observation period, has different expressions for G_{kn} as given by (2.16), the derivation of the new correlation model is more involved than in the case of the white noise excitation. To this end, in Appendix B, we compute and simplify $r_x(\tau)$ for each lag separately. By observing the patterns of the correlation functions for individual lags, a general expression for $r_x(\tau)$ can be obtained as

$$r_x(\tau) = \frac{1}{N} \sum_{k=1}^{P} \sum_{j=1}^{P} \eta_k \eta_j \Theta \, p_k^{\tau} \, , \quad \Theta = \Theta_I + \Theta_{II}, \tag{2.17}$$

$$\begin{aligned}
\Theta_I &= \left[\frac{1 - (p_k p_j)^T}{(1 - p_k p_j)(1 - p_k^T)(1 - p_j^T)} \right] \\
&\times \; [(\lambda - 1) - g(p_k) - g(p_j) + g(p_k p_j)] \\
&+ \left(\frac{1 - (p_k p_j)^{\Delta}}{1 - p_k p_j} \right) \left(\frac{1 - p_k^{\lambda T}}{1 - p_k^T} \right) \left(\frac{1 - p_j^{\lambda T}}{1 - p_j^T} \right), \\
\Theta_{II} &= \begin{cases} 0, & \tau = 0 \\ \dfrac{p_j^T \sum_{l=1}^{\tau} (p_k p_j)^{-l}}{1 - p_j^T} [(\lambda - 1) - g(p_j)], & \tau > 0 \end{cases}
\end{aligned}$$

In the above expressions, $g(\alpha)$ and Δ are given by

$$\begin{aligned}
g(\alpha) &= \alpha^T \left(\frac{1 - \alpha^{(\lambda-1)T}}{1 - \alpha^T} \right) \\
\Delta &= \begin{cases} 0, & R - \tau < 0 \\ R - \tau, & R - \tau \geq 0 \end{cases}
\end{aligned}$$

where $R \geq 0$ is the number of remaining samples after the last period and is given by

$$R = N - (\lambda - 1)T \tag{2.18}$$

In practical applications, the value of T is sufficiently large such that the expressions in $r_x(\tau)$ corresponding to the high powers of poles can be neglected. In such a case, the expression for $r_x(\tau)$ given by (2.17) can be simplified as

$$\begin{aligned} r_x(\tau) &= \sum_{k=1}^{P} \psi_{TF_k} p_k^{\tau},\ \tau = 0, 1, \ldots, M-1; M < \frac{T}{2}, \\ \psi_{TF_k} &= \frac{1}{N} \sum_{j=1}^{P} \eta_k \eta_j \Theta_f, \\ \Theta_f &= \begin{cases} \dfrac{\lambda - (p_k p_j)^{\Delta}}{1 - p_k p_j} = \Theta_{f_0}, & \tau = 0 \\ \Theta_{f_0} + (\lambda - 1) p_j^T \sum_{l=1}^{\tau} (p_k p_j)^{-l}, & \tau > 0 \end{cases} \end{aligned} \tag{2.19}$$

When the excitation is of infinite-duration, i.e., N tends to infinity, λ in (2.14) can be approximated as $\frac{N}{T}$; then in this case, (2.19) can be simplified as

$$\begin{aligned} r_x(\tau) &= \sum_{k=1}^{P} \psi_{TI_k} p_k^{\tau}, \tau = 0, 1, \ldots, M-1; M < \frac{T}{2}, \\ \psi_{TI_k} &= \frac{1}{T} \sum_{j=1}^{P} \eta_k \eta_j \Theta_i, \\ \Theta_i &= \begin{cases} \dfrac{1}{1 - p_k p_j} = \Theta_{i_0}, & \tau = 0 \\ \Theta_{i_0} + p_j^T \sum_{l=1}^{\tau} (p_k p_j)^{-l}, & \tau > 0 \end{cases} \end{aligned} \tag{2.20}$$

2.2.4 Estimation of AR Parameters of Noisy ARMA System Using ARMAC Model

Using a common notation ψ_k to represent ψ_{W_k}, ψ_{TF_k} or ψ_{TI_k}, the correlation models of an ARMA signal for the white noise and impulse-train excitations, as obtained in the last two subsections and given by (2.11), (2.19), and (2.20) can be expressed in a

unified manner given by

$$r_x(\tau) = \sum_{k=1}^{P} \psi_k p_k^{\tau},\ \tau = 0,\ 1, \ldots,\ M-1 \tag{2.21}$$

Although ψ_k may vary slightly as τ increases, within the range of M mentioned in Section 2.2.2 and 2.2.3, this variation is negligible and therefore, within this range it can be treated as a constant. Note that ψ_k can be real or complex depending on whether the poles are real or complex. As the ARMA signal under consideration is real, complex poles always appear in complex conjugate pairs. Letting $p_k = r_k e^{j\omega_k}$, $\psi_k = \zeta_k e^{j\upsilon_k}$, and θ be the number of real poles plus the number of pairs of complex conjugate poles, (2.21) can be rewritten as

$$r_x(\tau) = \sum_{l=1}^{\theta} r_l^{\tau}[\alpha_l \cos(\omega_l \tau) + \beta_l \sin(\omega_l \tau)], \tau = 0, 1, \ldots, M-1 \tag{2.22}$$

where $\alpha_l = \zeta_l \cos \upsilon_l$, $\beta_l = -\zeta_l \sin \upsilon_l$. Each of the θ terms in the summation of (2.22), namely $F_l(\tau) = r_l^{\tau}[\alpha_l \cos(\omega_l \tau) + \beta_l \sin(\omega_l \tau)]$, can be estimated sequentially from the M lags of $r_y(\tau)$, which is available from the noisy observations. In $r_y(\tau)$, the effect of noise is dominant mainly at $\tau = 0$. Hence, in order to reduce the noise effect, $\tau > 0$ is considered. The parameters r_l, ω_l, α_l and β_l of each component $F_l(\tau)$, with $\tau > 0$, are determined such that the total squared error between the $(l-1)$th residual function and $F_l(\tau)$ is minimized. We define the lth residual function as

$$\begin{aligned} \Re_l(\tau) &= \Re_{l-1}(\tau) - F_l(\tau);\ l = 1, 2, \ldots, \theta - 1; \\ \Re_0(\tau) &= r_y(\tau) \end{aligned} \tag{2.23}$$

Then, the total squared error for the minimization problem can be expressed as

$$J_l = \sum_{\tau=1}^{M-1} |\Re_{l-1}(\tau) - F_l(\tau)|^2,\ l = 1,\ 2,\ \ldots,\ \theta \tag{2.24}$$

Our objective is to find the optimum solution for r_l, ω_l, α_l and β_l using a search algorithm. In order to solve this optimization problem, we first equate the partial derivatives of J_l with respect to r_l and ω_l to zero, yielding

$$\begin{bmatrix} \sum_{\tau=1}^{M-1} r_l^{2\tau}\cos^2(\omega_l\tau) & \sum_{\tau=1}^{M-1} r_l{}^{2\tau} sc(\omega_l\tau) \\ \sum_{\tau=1}^{M-1} r_l{}^{2\tau} sc(\omega_l\tau) & \sum_{\tau=1}^{M-1} r_l{}^{2\tau}\sin^2(\omega_l\tau) \end{bmatrix} \begin{bmatrix} \alpha_l \\ \beta_l \end{bmatrix} = \begin{bmatrix} \sum_{\tau=1}^{M-1} \Re_{l-1}(\tau) r_l{}^{\tau}\cos(\omega_l\tau) \\ \sum_{\tau=1}^{M-1} \Re_{l-1}(\tau) r_l{}^{\tau}\sin(\omega_l\tau) \end{bmatrix} \tag{2.25}$$

where $sc(\omega_l\tau) = \sin(\omega_l\tau)\cos(\omega_l\tau)$. Now, we can formulate an optimization problem for minimizing J_l given by (2.24) with respect to r_l and ω_l subject to the constraint given by (2.25), and within some specified domain of values for r_l and ω_l in the range $0 < r_l < 1$ and $0 \leq \omega_l \leq \pi$, respectively. The values of $r_l = \hat{r}_l$ and $\omega_l = \hat{\omega}_l$ corresponding to the global minimum of J_l, are chosen as the estimate of the pole locations. If the estimated $\hat{\omega}_l$ has a value of 0 or π, this estimate represents a real pole. It is to be noted that the stability of the estimated ARMA model is guaranteed because of the constraint imposed on r_l in the process of minimization of J_l. Proceeding in this manner, all the P poles of the ARMA system can be estimated, and then, the AR parameters of the ARMA model can be determined easily by using (2.2).

The correlation model proposed in this section offers a two-fold advantage. Firstly, unlike the conventional recursive relation [33], the proposed model provides with a direct non-recursive relationship between the poles of the ARMA system and the ACF. Secondly, the proposed ARMAC model can be used in a correlation-fitting approach to estimate the AR parameters. The novelty of the proposed scheme for

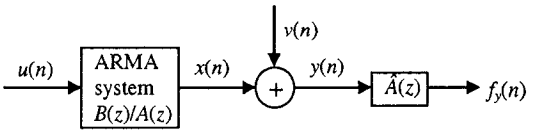


Figure 2.1: Filtering ARMA signal in noise by the estimated AR polynomial.

the AR parameter estimation lies in its ability to extract a noise-free correlation function from a noisy one through a model fitting-based approach instead of using the noisy one directly. Hence, the proposed technique can be expected to provide a more accurate AR parameter estimation even at a very low SNR.

2.3 Estimation of MA Parameters

2.3.1 Problems Statement

Identification of the zeros of the underlying ARMA system from the noise-corrupted observations is a difficult problem, since the additive noise directly affects the estimation of the system zeros. In a two-step ARMA identification method, a residual signal can be obtained by filtering the noisy ARMA signal $y(n)$ with the estimated AR polynomial $\hat{A}(z)$, giving

$$f_y(n) = y(n) + \sum_{k=1}^{P} \hat{a}_k y(n-k) \tag{2.26}$$

Fig. 2.1 shows the filtering process in the presence of noise $v(n)$. We assume that an accurate AR estimate of the ARMA system has already been obtained in the first step, i.e., $\hat{A}(z) \approx A(z)$. Then, (2.26) can be rewritten as

$$f_y(n) = \hat{B}(z)u(n) + \hat{A}(z)v(n) = f_x(n) + f_w(n) \tag{2.27}$$

where $f_x(n)$ represents the residual signal with respect to the noise-free ARMA signal $x(n)$ and $f_w(n)$ corresponds to the noisy part due to the additive white Gaussian

noise $v(n)$. It is evident that $f_w(n)$ is a colored MA(P) sequence, even if the original noise $v(n)$, which corrupts the ARMA system, is white. Thus, $f_y(n)$ can be treated as an MA sequence of order P. In this section, we propose a new algorithm that first eliminates the noise effect in the correlation domain and then determines the MA parameters from the inverse autocorrelation function (IACF). Similar to the problem of AR identification, both the white noise and impulse-train excitations are considered.

2.3.2 White Noise Excitation

The ACF of $f_y(n)$ can be obtained as

$$r_{f_y}(\tau) = E\left[f_y(n)f_y(n+\tau)\right] = r_{f_x}(\tau) + r_{f_w}(\tau) + r_{f_c}(\tau) \tag{2.28}$$

where $r_{f_x}(\tau)$ and $r_{f_w}(\tau)$ are, the ACFs of $f_x(n)$ and $f_w(n)$, respectively, and $r_{f_c}(\tau)$ represents the terms of cross-correlation between $f_x(n)$ and $f_w(n)$, which after some manipulation, can be expressed as

$$\begin{aligned} r_{f_c}(\tau) = r_{uv}(\tau) + r_{vu}(\tau) + \sum_{k=1}^{P} \hat{a}_k \{r_{uv}(\tau - k) + r_{vu}(\tau + k)\} \\ + \sum_{k=1}^{Q} \hat{b}_k \{r_{uv}(\tau + k) + r_{vu}(\tau - k)\} + \sum_{k=1}^{P}\sum_{l=1}^{Q} \hat{a}_k \hat{b}_l r'(\tau, l, k) \end{aligned} \tag{2.29}$$

with $r'(\tau, l, k) = r_{uv}(\tau - k + l) + r_{vu}(\tau + k - l)$. Note that the last term, $r_{f_c}(\tau)$, in (2.28) is expressed in terms of the cross-correlations between the excitation noise $u(n)$ and the additive noise $v(n)$. Due to the fact that both $u(n)$ and $v(n)$ are white noise, their cross-correlations can be neglected. As $f_x(n)$ can be treated as the output of the MA(Q) system $\hat{B}(z)$ excited by a white noise input $u(n)$ with variance σ_u^2, $r_{f_x}(\tau)$

can be expressed in terms of the MA parameters $\hat{b}_l$ as

$$r_{f_x}(\tau) = \begin{cases} \sigma_u^2 \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{2.30}$$

Similarly, $r_{f_w}(\tau)$ can be obtained as

$$r_{f_w}(\tau) = \begin{cases} \sigma_v^2 \sum_{l=0}^{P-\tau} \hat{a}_l \hat{a}_{l+\tau} & |\tau| \leq P \\ 0, & \text{otherwise} \end{cases} \tag{2.31}$$

Thus, using (2.30) and (2.31) in (2.28), $r_{f_y}(\tau)$ can be expressed as

$$r_{f_y}(\tau) = \begin{cases} \sigma_u^2 \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau} + \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, |\tau| \leq Q \\ \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, \ Q < |\tau| \leq P \\ 0, \ |\tau| > P \end{cases} \tag{2.32}$$

In order to evaluate the noise variance, we can use (2.32) for $Q < |\tau| \leq P$ provided that an estimate of $r_{f_y}(\tau)$, denoted as $\hat{r}_{f_y}(\tau)$, has already been obtained. Using (2.32) for $Q < |\tau| \leq P$, an estimate of the noise variance can be obtained as

$$\hat{\sigma}_v^2 = \hat{r}_{f_y}(\tau) / \textstyle\sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, \quad Q < |\tau| \leq P \tag{2.33}$$

Thus, with $\hat{\sigma}_v^2$ and $\hat{A}(z)$ already known, the noise-compensated $r_{f_y}(\tau)$ or an estimate of $r_{f_x}(\tau)$ can be obtained using (2.32) for $|\tau| \leq Q$. As seen from (2.30), the MA parameters $\{b_l\}$ can then be estimated from $\hat{r}_{f_x}(\tau)$. In order to compute $r_{f_y}(\tau)$, we employ the scheme proposed in [101], where P_a additional lags after $|\tau| = P$ have been used. Once the estimates $\hat{\sigma}_v^2$ and $\hat{r}_{f_y}(\tau)$ are determined, $r_{f_x}(\tau)$ can be estimated as

$$\hat{r}_{f_x}(\tau) = \begin{cases} \hat{r}_{f_y}(\tau) - \hat{\sigma}_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{2.34}$$

With the estimate $\hat{r}_{f_x}(\tau)$ computed, it is seen from (2.30) that the estimation of the MA parameters is a nonlinear problem. Recall that we have proposed in the previous

section an LS model-fitting based approach for the estimation of the AR parameters, which is capable of providing with a better estimate even in a heavy noisy condition. We now devise a scheme whereby the proposed model-fitting method of Section 2.2 can also be employed for the estimation of the MA parameters.

It is known that the inverse ACF (IACF) of an MA(Q) process is of the same form as the ACF of the corresponding AR(Q) process [102]. The IACF of $f_x(n)$ can be computed from the estimate $\hat{r}_{f_x}(\tau)$ as

$$\hat{\phi}_{f_x}(\tau) = F^{-1}\left[\frac{1}{\hat{\Phi}_{f_x}(e^{j\omega})}\right] = F^{-1}\left[\frac{1}{\sum_{m=-Q}^{Q}\hat{r}_{f_x}(m)e^{-j\omega m}}\right] \tag{2.35}$$

where F^{-1} denotes the operation of the inverse Fourier transform and $\hat{\Phi}_{f_x}(e^{j\omega})$ is the estimated PSD of the MA(Q) signal $f_x(n)$. Then, the IACF can be associated with the corresponding AR(Q) system by

$$\sum_{\tau=-\infty}^{\infty}\hat{\phi}_{f_x}(\tau)z^{-\tau} = \frac{1}{\hat{\Phi}_{f_x}(z)} = \frac{\hat{\sigma}_d^2}{\hat{B}(z)\hat{B}(z^{-1})} \tag{2.36}$$

where $\hat{\sigma}_d = \hat{\sigma}_u^{-1}$. It is seen that the IACF $\hat{\phi}_{f_x}(\tau)$ corresponds to the ACF of the AR(Q) system $\hat{B}(z)$ excited by the white noise $d(n)$ with variance $\hat{\sigma}_d^2$. Noting that the ARMA correlation model proposed in Section 2.2 is also valid for AR systems, the LS model-fitting based approach described in Section 2.2 can be readily applied to the estimation of the parameters $\{b_l\}$ of $B(z)$ from $\hat{\phi}_{f_x}(\tau)$.

Since the given system is assumed to be minimum-phase, in the above MA estimation, only a valid MA(Q) correlation sequence $\hat{r}_{f_x}(\tau)$ (i.e., the correlation sequence that gives exactly Q zeros inside the unit circle) should be used for the computation of the IACF. The validity of $\hat{r}_{f_x}(\tau)$ needs to be checked before computing $\hat{\phi}_{f_x}(\tau)$. This can be done easily by using the polynomial rooting described in [33]. It is clear from (2.34) that the estimation accuracy of the noise variance σ_v^2 would affect the validity of the estimated correlation sequence $\hat{r}_{f_x}(\tau)$. For a better estimation accuracy,

the values of $\hat{\sigma}_v^2$ resulting from the different choices of lags τ (see (2.33)) are averaged and used in (2.34). If the averaged noise variance estimate, denoted by $\widetilde{\hat{\sigma}}_v^2$, still does not produce a valid MA(Q) correlation sequence, a small neighborhood of $\widetilde{\hat{\sigma}}_v^2$ is searched to find a valid correlation sequence. Alternatively, validation schemes such as the over-parameterized algorithm of [101] can be used to obtain a valid correlation sequence.

2.3.3 Impulse-train Excitation

In the case of impulse-train excitation $u_i(n)$ with a period of T as given by (2.14), (2.27) can be rewritten as

$$f_{y_i}(n) = \hat{B}(z)u_i(n) + \hat{A}(z)v(n) = f_{x_i}(n) + f_w(n) \tag{2.37}$$

where $f_{x_i}(n)$ represents the residual signal with respect to the noise-free ARMA signal $x(n)$ and $f_w(n)$ corresponds to the noisy part due to the additive noise $v(n)$. Thus, $f_{x_i}(n)$ is the output of an MA(Q) system $\hat{B}(z)$ excited by the periodic impulse-train $u_i(n)$ and $f_w(n)$ is an MA(P) sequence excited by the white noise $v(n)$. Since the impulse response of an MA(Q) system, say $h_Q(n)$, vanishes beyond $n = Q$ samples, $f_{x_i}(n)$ is a periodic repetition of $h_Q(n)$ without overlaps when $T > Q$. The autocorrelation of $f_{y_i}(n)$ can be expressed as

$$r_{f_{y_i}}(\tau) = r_{f_{x_i}}(\tau) + r_{f_w}(\tau) \tag{2.38}$$

where $r_{f_{x_i}}(\tau)$ is the ACF of $f_{x_i}(n)$ and $r_{f_w}(\tau)$ is given by (2.31). Note that, similar to the case of white noise excitation (see (2.28)), the crosscorrelation terms between $f_{x_i}(n)$ and $f_w(n)$ in (2.38) has been neglected since, they consist of crosscorrelations between $u_i(n)$ and $v(n)$, which can be considered to be uncorrelated. The expression

for $r_{f_{x_i}}(\tau)$ can be derived as

$$r_{f_{x_i}}(\tau) = \sum_{k=-(\lambda-1)}^{(\lambda-1)} (\lambda - |k|)\, r_{h_Q}(\tau - kT) \tag{2.39}$$

$$r_{h_Q}(\tau) = \begin{cases} \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{2.40}$$

Recall that λ is the total number of impulses in the impulse-train excitation defined in (2.14). It is to be mentioned that $r_{f_{x_i}}(\tau)$ is periodic with period T, and therefore, in order to avoid aliasing in the correlation domain, one must choose $T > 2Q+1$. In the region $|\tau| < T - Q$, $r_{f_{x_i}}(\tau)$ reduces to

$$r_{f_{x_i T}}(\tau) = \begin{cases} \lambda \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau}, & |\tau| \leq Q \\ 0, & Q < |\tau| < T - Q \end{cases} \tag{2.41}$$

As seen from (2.38), the component $r_{f_{x_i}}(\tau)$ is a periodically repeated scaled version of $r_{h_Q}(\tau)$ (see (2.39)), and the component $r_{f_w}(\tau)$ vanishes beyond $\tau = P$ samples. Using (2.31) and (2.41), for $T > P + Q + 1$, (2.38) can be rewritten as

$$r_{f_{y_i}}(\tau) = \begin{cases} \lambda \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau} + \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, |\tau| \leq Q \\ \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, Q < |\tau| \leq P \\ r_{f_{x_i}}(\tau),\ |\tau| > P \end{cases} \tag{2.42}$$

Note that the format for $r_{f_{y_i}}(\tau)$ as given above is similar to that of $r_{f_y}(\tau)$ given by (2.32). It is seen from (2.39) and (2.42) that a non-zero value of $r_{f_{y_i}}(\tau)$ beyond $\tau = P$ appears only after $\tau = T - Q$. Hence, the number of additional lags (P_a) to be used in the computation of $r_{f_{y_i}}(\tau)$ is restricted to $P_a < T - (P + Q + 1)$. In practice, $T \gg P$ and thus, the condition $T > P + Q + 1$ will be automatically satisfied. Similar to the case of white noise excitation, an accurate estimation of $r_{f_{y_i}}(\tau)$ is first obtained and the estimate $\hat{r}_{f_{y_i}}(\tau)$ can then be used to calculate σ_v^2 according to the method

described at the end of Section 2.3.2. Finally an estimate of $r_{f_{x_i}}(\tau)$ can be obtained as

$$\hat{r}_{f_{x_i}}(\tau) = \begin{cases} \hat{r}_{f_{v_i}}(\tau) - \tilde{\hat{\sigma}}_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{2.43}$$

in which $\hat{r}_{f_{x_i}}(\tau)$ has been set to zero for $\tau > Q$. In the original expression for $r_{f_{x_i}}(\tau)$ as given by (2.39), when $r_{f_{x_i}}(\tau)$ is set to zero beyond $\tau = Q$, it can be written as

$$r_{f_{x_i}}(\tau) = \lambda \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau} = \lambda r_{h_Q}(\tau), |\tau| \leq Q \tag{2.44}$$

It is clear from (2.42), (2.43) and (2.44) that $\hat{r}_{f_{x_i}}(\tau)$ is the estimate of $r_{f_{x_i}}(\tau)$, which is a scaled version of $r_{h_Q}(\tau)$ given by (2.40). Hence, as in the case of white noise excitation (see (2.36)), the IACF $\hat{\phi}_{f_{x_i}}(\tau)$ computed from $\hat{r}_{f_{x_i}}(\tau)$, can be considered as the ACF of the $AR(Q)$ system $\hat{B}(z)$ excited by a white noise with variance $1/\lambda$. Therefore, the LS model-fitting based approach can be employed to determine the MA parameters from $\hat{\phi}_{f_{x_i}}(\tau)$, once the validity check for $\hat{r}_{f_{x_i}}(\tau)$ has been carried out.

The complete algorithm, whose development started in the previous section and completed in this one, will hereafter be referred to as ARMAC method. The main steps of this algorithm are summarized as Algorithm I.

Algorithm I: The Proposed ARMAC Method

1. Compute the autocorrelation $r_y(\tau)$ from noisy observation $y(n)$ using (2.5).

2. Determine the poles using the residue-based least-squares (RBLS) technique (Section 2.2.4) comprising the steps:

   i. Obtain the initial estimates of the pole locations from the peaks of the smoothed FFT of $y(n)$.

ii. For each value of l, calculate J_l using (2.23)−(2.25) and determine the desired pole p_l corresponding to the global minimum value of J_l.

iii. Repeat (ii) until all the P poles are obtained.

iv. Compute the AR parameters from the estimated poles.

3. Determine the residual signal $f_y(n)$ by filtering $y(n)$ using the estimated AR polynomial obtained in Step 2 and estimate its autocorrelation $r_{f_y}(\tau)$.

4. Obtain the estimate of the noise variance σ_v^2 using (2.33) and the noise-compensated ACF $\hat{r}_{f_x}(\tau)$ of the residual signal using (2.34) and the method described in Section 2.3.2.

5. Compute IACF $\hat{\phi}_{f_x}(\tau)$ corresponding to $\hat{r}_{f_x}(\tau)$ using (2.35).

6. Determine zeros from $\hat{\phi}_{f_x}(\tau)$ using Step 2 and obtain the required MA parameters from estimated zeros.

2.4 Simulation Results

In this section, we perform a number of simulations for the identification of ARMA systems in the presence of noise. We consider synthetic ARMA signals as well as natural speech signals corrupted by additive noise. First, in order to show the accuracy of the proposed ARMAC model, we compute the correlation function using the proposed model and the conventional correlation estimator given by (2.5). The purpose here is to show that the values of the correlation function computed with the known parameter values of the proposed model and the conventional estimator that uses only the system output are reasonably close. For the proposed models we use

(2.11) for the white noise excitation and (2.19) and (2.20) for the impulse-train excitation. In this simulation, an ARMA(4, 3) system with parameters as given in Table 2.1 is used. We have considered $N = 512$ as an example for smaller data length [79] and $N = 10,000$ for a relatively large data length. For the purpose of comparison, the correlation sequences are normalized to unity and we have chosen $M = 40$ and $T = 94$. The correlation results using the conventional estimator and the proposed models are shown in Fig. 2.2. The close match between the two correlation functions obtained by the conventional estimator and the proposed model clearly shows the accuracy of the proposed model. Next, the performance in terms of the accuracy and consistency of the estimated parameters of the proposed method is obtained and compared with that of the conditional maximum-likelihood (ML) method [34], also referred to as the prediction error method, the ARMA cepstrum recursion (ACR) method [45], and a method (abbreviated as DYW method) in which the Durbin's scheme for the MA parameter estimation is combined with the least-squares modified Yule-Walker (LSMYW) equations for the AR parameter estimation [33]. We have also considered the autocorrelation matching (ACM) method proposed in [77] and comparative results are shown for some systems. In the ACM method the AR parameters are estimated from the ACF of noisy observations and it also estimates the MA parameters from the ACF using an autocorrelation matching technique under the assumption that the excitation power is known.

2.4.1 Results on Synthetic ARMA Systems

(a) White Noise Excitation: A noisy ARMA signal is generated according to (2.1) and (2.6) with $N = 4,000$ and $\sigma_u^2 = 1$, where the variance of the white Gaussian noise

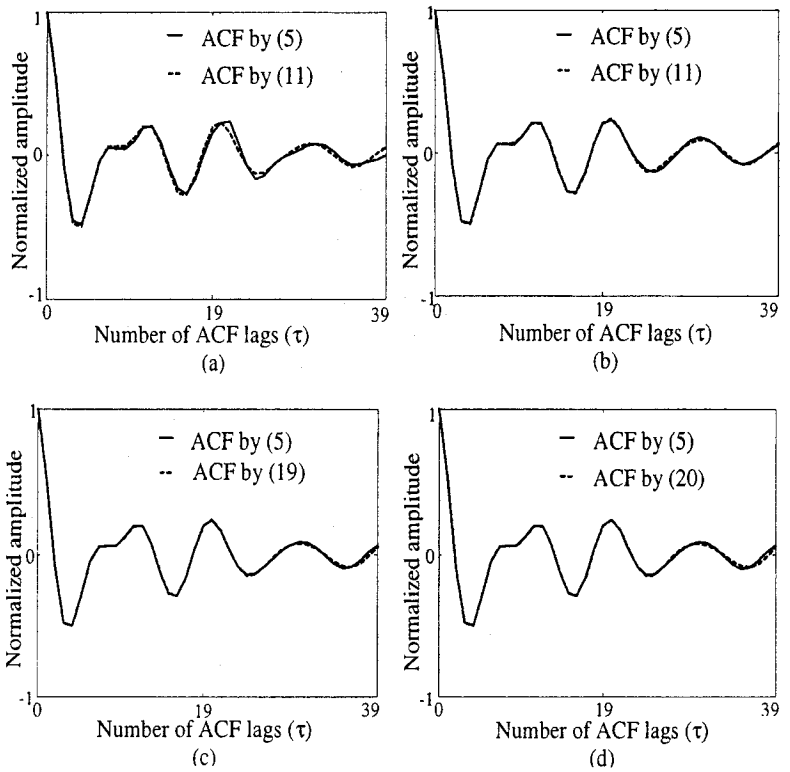


Figure 2.2: ACFs generated by using the conventional estimator (2.5) and the proposed ARMAC model for the white noise excitation with (a) $N = 512$ and (b) $N = 10,000$, and for the impulse-train excitation with (c) $N = 512$ and (d) $N = 10,000$.

can be appropriately set according to the desired SNR defined as

$$\text{SNR} = 10\log_{10}[P_x/P_v] \ \text{dB} \tag{2.45}$$

where $P_x = \sum_{n=0}^{N-1} x(n)^2)$ and $P_v = \sum_{n=0}^{N-1} v(n)^2$. The ARMAC model parameters are then determined using the RBLS optimization technique described in Section 2.2. In our simulation, the search range for r_l is chosen in the range $0.5 \leq r_l \leq 0.99$, that allows the identification of even those systems whose autocorrelations decay very fast. The initial estimates of ω_l are chosen from the location of the peaks of the

smoothed FFT [33] of the noisy signal $y(n)$. The search range for ω_l is chosen as 0.1π 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. The number of lags for the ACF is set to be $M = 10P$. In the estimation of the MA parameters, $P_a = 60$ additional lags of $\hat{r}_{f_y}(\tau)$ is used. In the case, $\widetilde{\hat{\sigma}}_v^2$ cannot produce a valid MA(Q) correlation sequence, as described in Section 2.3, a neighborhood search, say in the region of 5% of $\widetilde{\hat{\sigma}}_v^2$ with a resolution of 0.01 is performed. As a matter of fact, in our extensive simulations, it has been observed that in most of the cases, less than 10 searches are sufficient to obtain a valid correlation sequence. Next, the IACF is computed using the N_f-point FFT and IFFT operation, where N_f is chosen as the smallest power of 2 greater than or equal to N. As mentioned in Section 2.3, the RBLS optimization technique is used to extract the parameters b_l of $B(z)$. In this case, the number of lags for the ACF is set to be $M = 10Q$ and an initial estimate is obtained from the smoothed FFT of the IACF. Other search parameters, such as the search range and the resolution, are kept the same as used above for the AR case. Each experiment contains $N_T = 100$ independent trials. We have conducted the experiments for the noisy cases, where the SNR varies from -10 dB to 15 dB at steps of 2.5 dB. The performance measurement criteria considered in our simulation study are (1) estimation mean, (2) the standard deviation from the mean (SDM), (3) the standard deviation from the given value (or the true value) (SDT), and (4) the average sum-squared error (ASSE) given by

$$\text{ASSE} = \frac{1}{N_T(P+Q)} \sum_{m=1}^{N_T} (E_a + E_b) \tag{2.46}$$

where $E_a = \sum_{k=1}^{P} [\hat{a}_k(m) - a_k]^2$, $E_b = \sum_{j=1}^{Q} [\hat{b}_j(m) - b_j]^2$, and $\hat{a}_k(m)$ and $\hat{b}_j(m)$ are the estimated parameters at the mth trial, and a_k and b_j the true values of the parameters.

Table 2.1: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited ARMA(4, 3) system at SNR = 10 dB

Methods	Estimated parameters							ASSE
	a_1	a_2	a_3	a_4	b_1	b_2	b_3	(dB)
ARMAC	−2.5745	3.3385	−2.2372	0.7540	−1.9909	1.7250	−0.6046	−22.78
(Proposed)	(±0.0473)	(±0.0961)	(±0.0849)	(±0.0356)	(±0.0654)	(±0.0716)	(±0.0660)	
	(±0.0515)	(±0.0961)	(±0.0927)	(±0.0424)	(±0.0704)	(±0.0716)	(±0.0675)	
ARMAX	−1.7592	2.0615	−1.3021	0.5421	−1.2484	1.0844	−0.5688	1.04
	(±0.8973)	(±1.3002)	(±0.8527)	(±0.1635)	(±0.8914)	(±0.8092)	(±0.1218)	
	(±1.2263)	(±1.8228)	(±1.2383)	(±0.2498)	(±1.1770)	(±1.0301)	(±0.1237)	
ACR	−2.2888	2.6758	−1.6265	0.5336	−1.8076	1.5132	−0.6678	−7.98
	(±0.0877)	(±0.1963)	(±0.1730)	(±0.0631)	(±0.0887)	(±0.1563)	(±0.0867)	
	(±0.3185)	(±0.6917)	(±0.5990)	(±0.2072)	(±0.2274)	(±0.2607)	(±0.1162)	
DYW	−2.2888	2.6758	−1.6265	0.5336	−1.5694	1.1192	−0.3351	−5.09
	(±0.0877)	(±0.1963)	(±0.1730)	(±0.0631)	(±0.4183)	(±0.5454)	(±0.2914)	
	(±0.3185)	(±0.6917)	(±0.5990)	(±0.2072)	(±0.6127)	(±0.8127)	(±0.3874)	
True	−2.5950	3.3390	−2.2000	0.7310	−2.0170	1.7218	−0.5904	

Table 2.2: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited ARMA(4, 3) system at SNR = −5 dB

Methods	Estimated parameters							ASSE
	a_1	a_2	a_3	a_4	b_1	b_2	b_3	(dB)
ARMAC	−2.5691	3.3268	−2.2216	0.7444	−1.9656	1.7066	−0.5632	−15.26
(Proposed)	(±0.0982)	(±0.1979)	(±0.1836)	(±0.0827)	(±0.2314)	(±0.2309)	(±0.0842)	
	(±0.1016)	(±0.1983)	(±0.1849)	(±0.0838)	(±0.2371)	(±0.2314)	(±0.0884)	
ARMAX	−0.9604	0.0950	−0.3906	−0.0352	−0.8425	−0.0273	0.3144	5.88
	(±0.5135)	(±0.5893)	(±0.3419)	(±0.0591)	(±0.5164)	(±0.5411)	(±0.2917)	
	(±1.7133)	(±3.2971)	(±2.6131)	(±0.7685)	(±1.2830)	(±1.8308)	(±0.9507)	
ACR	−0.6151	0.0435	0.1732	0.1717	−0.5020	−0.0336	0.0824	5.81
	(±0.1199)	(±0.1623)	(±0.1430)	(±0.0925)	(±0.1204)	(±0.1492)	(±0.1286)	
	(±1.9835)	(±3.2995)	(±2.3775)	(±0.5669)	(±1.5198)	(±1.7616)	(±0.6850)	
DYW	−0.6151	0.0435	−0.1732	0.1717	−0.4627	0.0312	0.0869	5.87
	(±0.1199)	(±0.1623)	(±0.1430)	(±0.0925)	(±0.5043)	(±0.4049)	(±0.2994)	
	(±1.9835)	(±3.2995)	(±2.3775)	(±0.5669)	(±1.6341)	(±1.7384)	(±0.7405)	
True	−2.5950	3.3390	−2.2000	0.7310	−2.0170	1.7218	−0.5904	

Tables 2.1 and 2.2 show the estimation results for the ARMA(4, 3) system where the SNR values are set to 10 dB and −5 dB, respectively. Tables 2.3 and 2.4 provide the corresponding estimation results for the ARMA(3, 2) system. Different ARMA systems are investigated in order to cover a wide range of possible combinations of pole-zero locations as well as types (i.e., real or complex conjugate). In the ARMA(3, 2) and ARMA(4, 3) systems the zeros and poles are closely spaced in order to demonstrate the estimation performance of the proposed scheme in dealing with

Table 2.3: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited ARMA(3, 2) system at SNR = 10 dB

Methods	Estimated parameters					ASSE
	a_1	a_2	a_3	b_1	b_2	(dB)
ARMAC	−2.2898	2.0340	−0.7229	−1.4843	0.6053	−26.86
(Proposed)	(±0.0286)	(±0.0523)	(±0.0277)	(±0.0405)	(±0.0399)	
	(±0.0288)	(±0.0525)	(±0.0300)	(±0.0551)	(±0.0528)	
ARMAX	−2.2685	1.9479	−0.6503	−1.6087	0.7196	−20.73
	(±0.0172)	(±0.0268)	(±0.0124)	(±0.0191)	(±0.0180)	
	(±0.0300)	(±0.0851)	(±0.0624)	(±0.0891)	(±0.0816)	
ACR	−2.2569	1.9749	−0.6880	−1.6289	0.8594	−18.35
	(±0.0298)	(±0.0506)	(±0.0228)	(±0.0394)	(±0.0561)	
	(±0.0468)	(±0.0739)	(±0.0327)	(±0.1143)	(±0.2265)	
DYW	(−2.2569	1.9749	−0.6880	−1.4187	0.5920	−13.14
	(±0.0298)	(±0.0506)	(±0.0228)	(±0.3593)	(±0.3038)	
	(±0.0468)	(±0.0739)	(±0.0327)	(±0.3737)	(±0.3076)	
True	−2.2930	2.0287	−0.7115	−1.5217	0.6400	

Table 2.4: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited ARMA(3, 2) system at SNR = −5 dB

Methods	Estimated parameters					ASSE
	a_1	a_2	a_3	b_1	b_2	(dB)
ARMAC	−2.2529	1.9826	−0.6858	−1.4449	0.6112	−17.45
(Proposed)	(±0.1059)	(±0.1517)	(±0.0979)	(±0.1317)	(±0.1350)	
	(±0.1133)	(±0.1585)	(±0.1012)	(±0.1525)	(±0.1380)	
ARMAX	−1.0125	0.6472	−0.0442	−0.8593	0.5524	0.33
	(±0.5493)	(±0.4505)	(±0.0570)	(±0.5504)	(±0.3851)	
	(±1.3934)	(±1.4531)	(±0.6697)	(±0.8612)	(±0.3949)	
ACR	−1.0902	0.2907	0.0283	0.9490	0.1985	0.67
	(±0.2145)	(±0.3187)	(±0.1445)	(±0.2178)	(±0.2909)	
	(±1.2218)	(±1.7670)	(±0.7538)	(±0.6127)	(±0.5287)	
DYW	−1.0902	0.2907	0.0283	−0.7582	0.1555	1.09
	(±0.2145)	(±0.3187)	(±0.1445)	(±0.5263)	(±0.4005)	
	(±1.2218)	(±1.7670)	(±0.7538)	(±0.9273)	(±0.6286)	
True	−2.2930	2.0287	−0.7115	−1.5217	0.6400	

such a difficult identification situation. In each table, the last row gives the true parameter values. The estimated values of the corresponding parameters obtained from the proposed and the three other methods are given in the preceding four rows. The corresponding values for the SDM and SDT are shown within the parentheses below the each estimated parameter value. The last column of each table provides the ASSE in dB. It is to be mentioned that to implement the ML method the '*armax*' command from the MATLAB System Identification Toolbox is used [34]. In the DYW method,

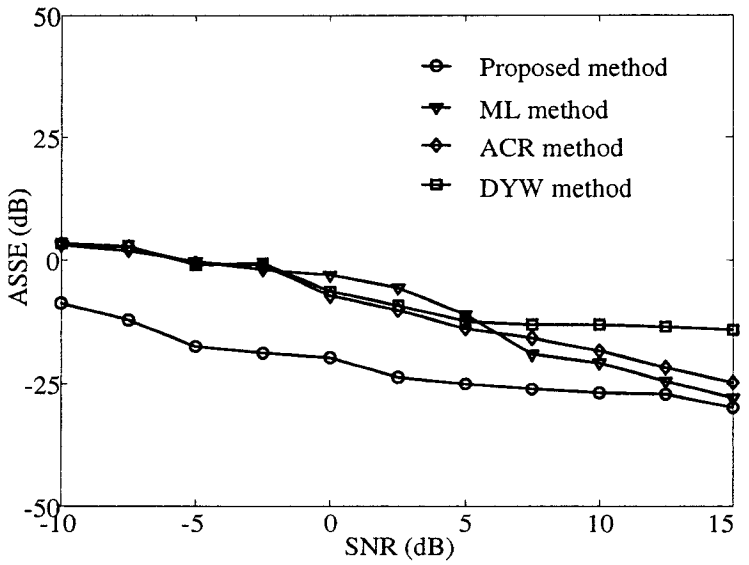


Figure 2.3: Effect of noise on the estimation accuracy for white noise excited ARMA(3, 2) system.

the LSMYW equations are used for the AR as well as the intermediate AR parameter estimation [33]. It is seen from these tables that the proposed method exhibits superior performance with respect to all the four performance indices at both levels of SNR. Very small values of SDM and SDT obtained from the proposed technique indicate a high degree of estimation consistency and accuracy. It is seen from Tables 2.1 and 2.3 that at SNR = 10 dB, although some of the other methods provide an acceptable performances, the estimation accuracy achieved by the proposed method is much higher. Tables 2.2 and 2.4 show that even at SNR = -5 dB, when the other methods fail to identify the system, the proposed method successfully estimates the parameters with sufficient accuracy.

Fig. 2.3 shows the ASSE values as a function of SNR levels for the four methods for the ARMA(3, 2) system with the true parameters given in Table 2.3. It is observed

from Fig. 2.3 that the ML and the ACR methods give estimation accuracy comparable to that provided by the proposed ARMAC method for SNR levels above 10 dB. However, the proposed method performs much better for levels of SNR as low as -5 dB.

Fig. 2.4 depicts the superimposed plots of the estimated poles and zeros from 25 realizations obtained by using the different methods at SNR $= -5$ dB for the ARMA$(5, 4)$ system whose parameters are given as

$$\begin{aligned} a_k &= \{1, -2.0825, 2.267, -2.1997, 1.8563, -0.811\} \\ b_j &= \{1, -1.6379, 1.5279, -1.2989, 0.6281\} \end{aligned}$$

For the purpose of comparison, the true poles and zeros are also plotted. Clearly, the estimated values obtained by using the proposed ARMAC method are much less scattered around the true values indicating a very high estimation accuracy in comparison to that achieved by the other methods.

The effect of data length on the estimation accuracy at different levels of SNR has also been investigated. In Fig. 2.5, the effect of data length on the ASSE (dB) at SNR $= 10$ dB and -5 dB for the same ARMA$(3, 2)$ system as the one used in Fig. 2.3 is shown. An important observation that can be made from this figure is that the proposed ARMAC method provides quite an accurate estimation results even at a data length as small as $N = 500$ samples. Similar to the other methods, the proposed method shows consistency of the results throughout most of the data length.

In order to show the effectiveness of the proposed technique, we carry out another comparison by providing the estimation results obtained by using the ARMAC and ACM [77] methods in Tables 2.5. For this purpose, an ARMA$(3, 2)$ system used in [77] is considered. As seen from the table, the proposed ARMAC method gives a

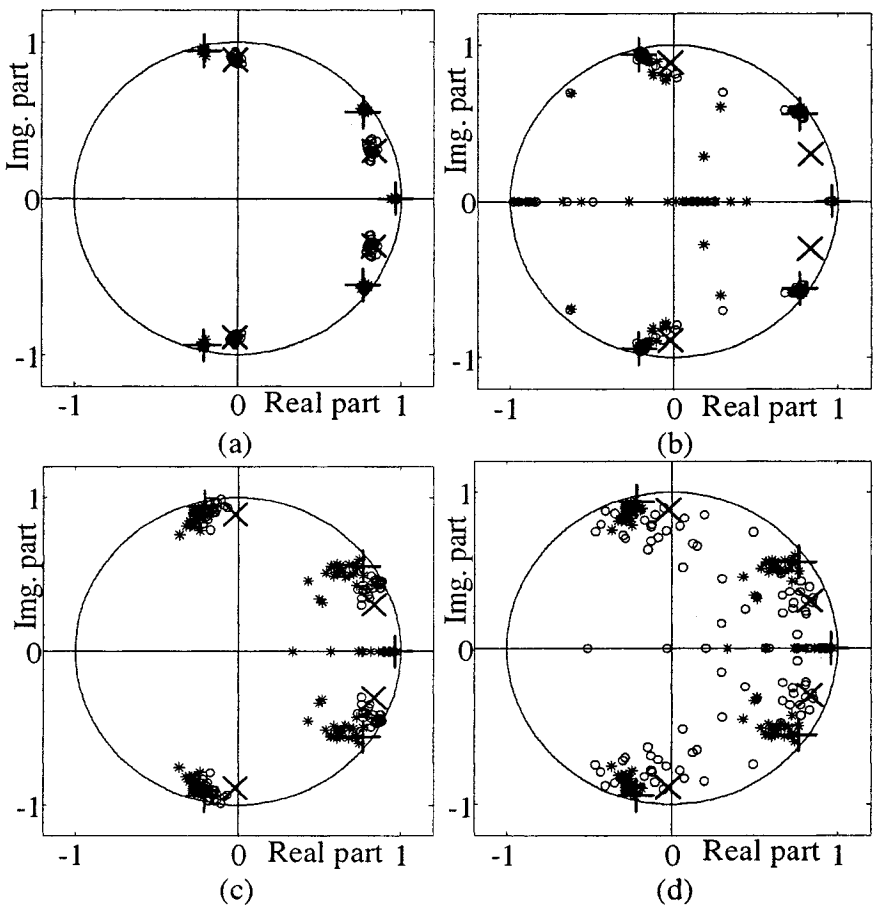


Figure 2.4: Superimposed pole-zero plot of ARMA(5, 4) system at SNR = −5 dB. × : true poles, + : true zeros, ∗ : estimated poles, ∘ : estimated zeros. (a) ARMAC (proposed), (b) ARMAX , (c) ACR, and (d) DYW method.

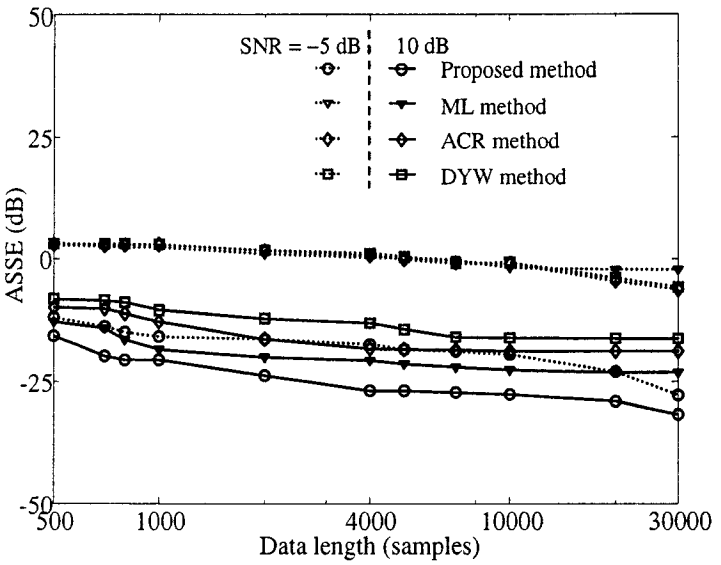


Figure 2.5: Effect of data length on the estimation accuracy for white noise excited ARMA$(3,2)$ system.

lower estimation variance even at a very low SNR of -5 dB. The estimation accuracy of the ACM method is affected in the presence of heavy noise due to its requirement of using high order models equivalent to the noisy ARMA process as well as due to the employment of the lattice filter based method.

(b) Impulse-Train Excitation:

We have also considered the problem of ARMA system identification with the periodic impulse-train excitations of different periods for various levels of noise. An impulse-train is generated using (2.14) with a known value of T. A noisy ARMA signal is generated according to (2.1) and (2.6) with $N = 4,000$. The simulations are run over $N_T = 100$ independent trials and the results averaged.

Tables 2.6 and 2.7 provide the estimation results for the impulse-train excited

Table 2.5: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited another ARMA(3, 2) system

	True values [77]	SNR=−5 dB		SNR=0 dB	
		ARMAC (Proposed)	ACM [77]	ARMAC (Proposed)	ACM [77]
a_1	−2.2990	−2.3059 (±0.0299) (±0.0307)	−2.2694 (±0.1628) (±0.1639)	−2.2998 (±0.0234) (±0.0234)	−2.3329 (±0.0269) (±0.0433)
a_2	2.1262	2.1133 (±0.0454) (±0.0472)	2.0521 (±0.3703) (±0.3776)	2.1186 (±0.0310) (±0.0327)	2.1334 (±0.0434) (±0.0440)
a_3	−0.7604	−0.7423 (±0.0292) (±0.0405)	−0.7219 (±0.3260) (±0.3266)	−0.7589 (±0.0206) (±0.0266)	−0.7447 (±0.0350) (±0.0357)
b_1	−0.8700	−0.8351 (±0.1770) (±0.1867)	−0.9866 (±0.5441) (±0.5664)	−0.8410 (±0.1072) (±0.1187)	−0.9708 (±0.5485) (±0.5577)
b_2	0.9200	0.8903 (±0.1310) (±0.1344)	0.7673 (±0.1755) (±0.2263)	0.9080 (±0.0427) (±0.0443)	0.8235 (±0.0873) (±0.1301)
	ASSE(dB)	−16.58	−8.95	−18.81	−11.77

ARMA(4, 3) system with $T = 70$ at SNR levels of 10 dB and −5 dB, respectively. It is seen from these tables that the proposed method provides quite an accurate estimation of the ARMA parameters, whereas the other methods are unable to identify the system at SNR = −5 dB and their performance is relatively poor at SNR = 10 dB.

The ASSE resulting from using various methods under the impulse-train excitation with $T = 70$ for the estimation of the same ARMA(3, 2) system as the one considered for the white noise excitation is shown in Fig. 2.6. As seen, the proposed ARMAC method provides a significantly better performance even at a very low SNR, whereas the performance of other methods deteriorates at SNR level below 10 dB.

Fig. 2.7 shows the effect of excitation period (T) on the estimation accuracy for the ARMA(4, 3) system used in Table 2.1 at SNR levels of 10 dB and −5 dB. It is seen from this figure that at SNR = −5 dB, all other methods give an ASSE value greater than 0 dB, whereas the ASSE of the proposed method is smaller even at SNR = −5 dB than that of the other method at SNR = 10 dB.

Table 2.6: Estimated parameters along with standard deviations (SDM and SDT) for impulse-train excited ARMA(4, 3) system at SNR = 10 dB

Methods	Estimated parameters							ASSE
	a_1	a_2	a_3	a_4	b_1	b_2	b_3	(dB)
ARMAC	−2.5924	3.3733	−2.2670	0.7772	−1.9767	1.6819	−0.5376	−24.09
(Proposed)	(±0.0088)	(±0.0209)	(±0.0198)	(±0.0076)	(±0.0626)	(±0.0748)	(±0.0544)	
	(±0.0092)	(±0.0402)	(±0.0699)	(±0.0468)	(±0.0744)	(±0.0848)	(±0.0758)	
ARMAX	−1.8613	2.1511	−1.3227	0.5088	−1.3528	1.1452	−0.5523	−0.28
	(±0.6839)	(±1.0494)	(±0.7195)	(±0.1514)	(±0.6801)	(±0.6577)	(±0.0996)	
	(±1.0030)	(±1.5851)	(±1.1346)	(±0.2688)	(±0.9507)	(±0.8746)	(±0.1066)	
ACR	−2.4617	3.0202	−1.9078	0.6302	−1.7292	1.5352	−0.6718	−12.54
	(±0.0560)	(±0.1302)	(±0.1172)	(±0.0417)	(±0.0621)	(±0.0904)	(±0.0556)	
	(±0.1446)	(±0.3443)	(±0.3148)	(±0.1091)	(±0.2944)	(±0.2073)	(±0.0986)	
DYW	(−2.2888	2.6758	−1.6265	0.5336	−1.9233	1.6487	−0.5098	−11.57
	(±0.0877)	(±0.1963)	(±0.1730)	(±0.0631)	(±0.2390)	(±0.3400)	(±0.2088)	
	(±0.3185)	(±0.6917)	(±0.5990)	(±0.2072)	(±0.2567)	(±0.3478)	(±0.2238)	
True	−2.5950	3.3390	−2.2000	0.7310	−2.0170	1.7218	−0.5904	

Table 2.7: Estimated parameters along with standard deviations (SDM and SDT) for impulse-train excited ARMA(4, 3) system at SNR = −5 dB

Methods	Estimated parameters							ASSE
	a_1	a_2	a_3	a_4	b_1	b_2	b_3	(dB)
ARMAC	−2.6121	3.3956	−2.2730	0.7691	−1.9908	1.6703	−0.5178	−15.76
(Proposed)	(±0.0804)	(±0.1566)	(±0.1534)	(±0.0762)	(±0.2297)	(±0.2217)	(±0.0660)	
	(±0.0822)	(±0.1665)	(±0.1699)	(±0.0852)	(±0.2312)	(±0.2276)	(±0.0981)	
ARMAX	−1.0803	0.1863	−0.4200	−0.0336	−0.9809	0.0720	0.3831	5.86
	(±0.5686)	(±0.8792)	(±0.5690)	(±0.0638)	(±0.5724)	(±0.8272)	(±0.4796)	
	(±1.6179)	(±3.2730)	(±2.6810)	(±0.7673)	(±1.1836)	(±1.8456)	(±1.0852)	
ACR	(−0.8804	0.2321	0.1681	0.1032	−0.8566	0.2047	0.1535	5.52
	(±0.1862)	(±0.2821)	(±0.2357)	(±0.1232)	±0.1865)	(±0.2792)	(±0.2303)	
	(±1.7247)	(±3.1197)	(±2.3798)	(±0.6398)	±1.1753)	(±1.5425	(±0.7787)	
DYW	(−0.8804	0.2321	0.1681	0.1032	−0.4830	0.0272	−0.0863	5.63
	(±0.1862)	(±0.2821)	(±0.2357)	(±0.1232)	±0.4861)	(±0.3943)	(±0.2441)	
	(±1.7247)	(±3.1197)	(±2.3798)	(±0.6398)	±1.6092)	(±1.7398)	(±0.5601)	
True	−2.5950	3.3390	−2.2000	0.7310	−2.0170	1.7218	−0.5904	

2.4.2 An Application for Vocal-tract System Identification

As an application of the proposed method, the identification of a vocal-tract system is performed using natural speech signals. Since, in this case, the true system parameters are not known, for the purpose of evaluating the estimation accuracy, non-parametric PSD is used. In order to estimate the vocal-tract system parameters, some English nasal sounds (voiced phonemes) from the TIMIT standard database with a sampling frequency of 16 KHz are used as the noise-free output observations. Instances of the

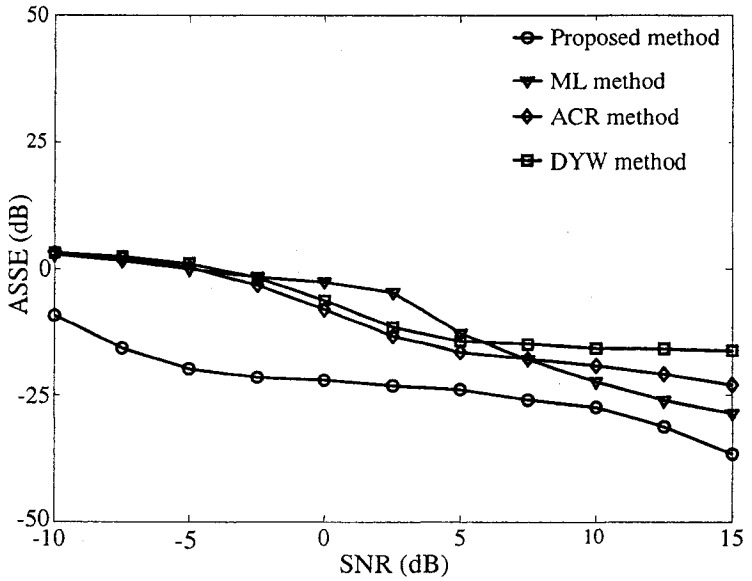


Figure 2.6: Effect of noise on the estimation accuracy for impulse-train excited ARMA(3, 2) system.

phonemes are extracted from the database according to the given transcriptions. No pre-filtering is performed in order to observe the accuracy of the pole-zero estimation over the entire range of frequency. With the estimated parameters of the vocal-tract considered as an ARMA system and the pitch-period, a speech phoneme can be synthesized using a value of the vocal-tract filter gain appropriately determined based on the RMS power level and the peak PSD of the natural speech frames [7]. In order to verify the estimation accuracy, the PSD of the synthesized speech is compared with that of the noise-free natural speech. It is to be mentioned that the synthesized sounds obtained by different methods were also played back in order to test the subjective quality.

Fig. 2.8 shows a comparison of the PSDs of the vocal-tract system obtained from different methods in noisy environments with respect to noise-free PSD. Considering

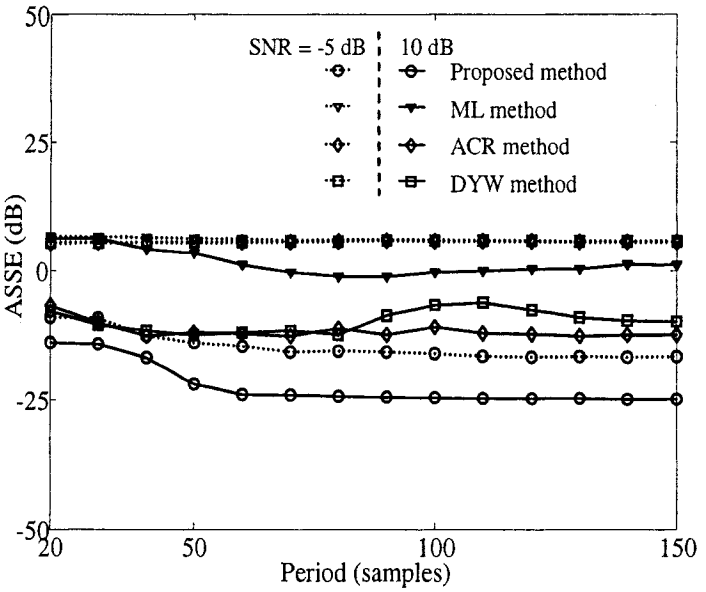


Figure 2.7: Effect of variation of excitation period (T) on the estimation accuracy for ARMA(4, 3) system.

the fact that the choice of the order of the vocal-tract filter depends on the characteristic of the specific phoneme, an ARMA$(12, 4)$ model is used for a naturally spoken nasal sound /m/ of the word '*him*' uttered by a female speaker. The estimated pole-zero locations of the vocal-tract system are averaged over 25 independent realizations of noisy environments and used to obtain the synthesized speech.

In our experiments, we consider the ACF lags in the range of 0 to $T/2$, where the excitation period or pitch (T) is estimated using the scheme of [103]. According to the general behavior of the vocal tract parameter, r_l is searched in the range $[0.8, 0.99]$ [93]. The search range for ω_l can be narrowed down based on the knowledge of the pole-zero locations of a particular phoneme [7], [93]. Fig. 2.8 shows the PSD of the synthesized speech using the various methods along with that of noise-free speech. In order to have a better understanding of the level of noise, the PSD of one of the 25 noisy

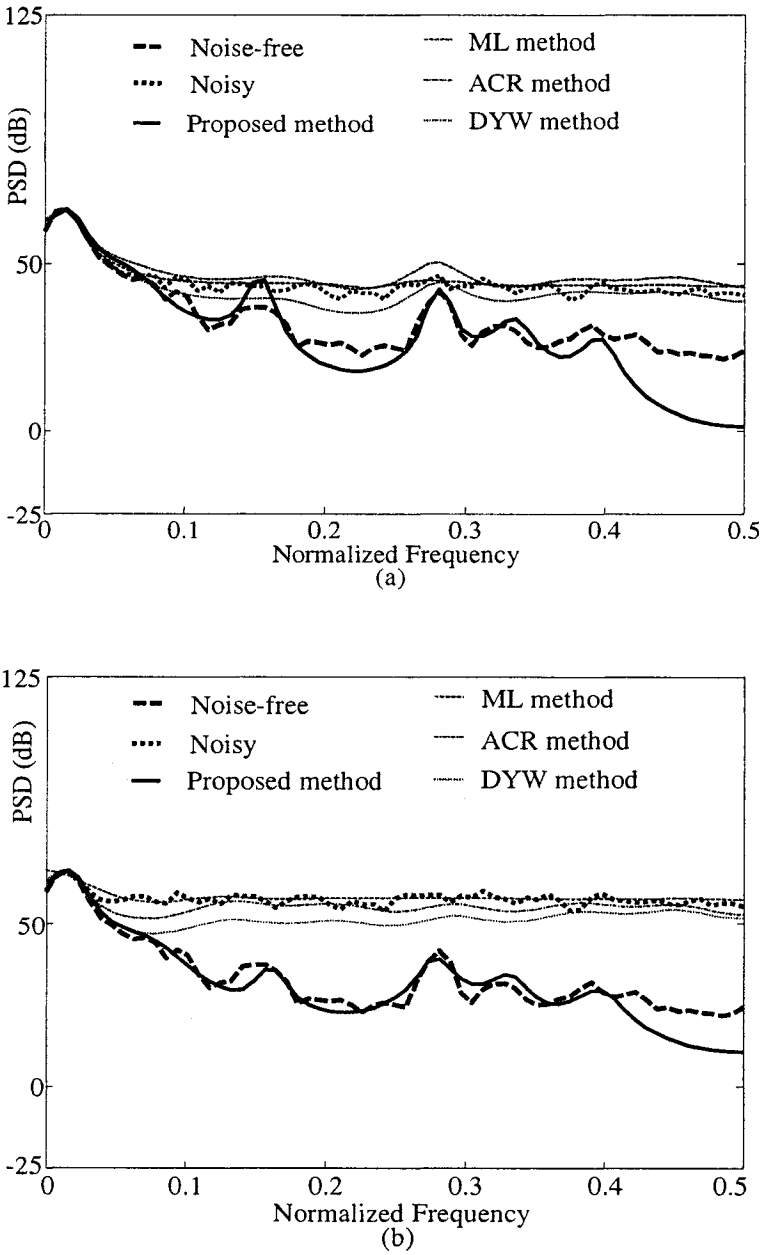


Figure 2.8: Spectrum comparison for the speech phoneme /m/ taken from an utterance '*him*' under noisy conditions at SNR levels of (a) 10 dB and (b) -5 dB.

signals is also included in the same figure. 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. On the other hand, the PSDs resulting from the other methods simply follow the PSD of the corresponding noisy observations, and thus, fail to estimate the parameters of the vocal-tract system. Quality of the synthesized sounds obtained by the proposed method under such a noisy condition were found far superior than that obtained by the other methods. The proposed ARMAC method provides a satisfactory identification performance of the vocal-tract system even at a very low SNR level of -5 dB.

2.5 Conclusion

In this Chapter, an effective methodology for the identification of ARMA systems using the observed output signal in the presence of heavy additive noise has been presented. The proposed method is based on developing a simple but accurate model, called the ARMAC model, in terms of the poles of the ARMA system. A significant feature of the new model is that it has been presented in a unified form in terms of the poles of the ARMA systems for both white noise and periodic impulse-train excitations of finite or infinite durations.

A residue-based least-squares (RBLS) correlation fitting optimization scheme that employs the ARMAC model has been presented for the estimation of the AR parameters. The proposed RBLS scheme has an advantage in the sense that it provides the flexibility of incorporating some *a priori* knowledge of the parameters, if available, to the process of parameter estimation. Since the presence of the noise in the residual signal in a two-stage ARMA system identification makes the problem of MA parameter estimation difficult, an efficient scheme using the estimated AR parameters and

output noise variance has been presented in order to reduce the effect of the noise in the correlation function of the residual signal. Then, for the identification of the MA part of the ARMA system, the estimation task has been transformed into a problem of correlation fitting of the inverse of the noise-compensated correlation function and solved again by employing the proposed RBLS scheme.

One of the assumptions made in the derivation of the proposed method is that the order of the ARMA model is known as done in most of the existing techniques [13], [43]–[45], [75]–[79], [96] for the ARMA system identification. In addition to this assumption of known ARMA(P, Q) model order with $P > Q$, it is assumed that the system has minimum phase and only the first-order poles and zeros. Eventhough these assumptions, as made in many other similar works in the literature, have been used in this paper to simplify the theoretical development of the proposed technique, they in fact do not impose undesirable limitations in dealing with most of the real-life situations. In the case of real-life data (i.e., the data in the presence of noise) a bad choice of the model order would affect the estimation accuracy of the parameters for all the methods. However, in the proposed algorithm, the AR estimation part works independent of the MA estimation part, and the poles are determined successively, a simple one or a pair of complex ones at a time. Thus, the accuracy of at least those poles which are estimated using the proposed method is not affected if the model order is chosen wrongly as (P', Q') for an ARMA(P, Q) model for the case when $Q' < P' < P$.

In order to investigate the estimation performance of the proposed method for systems with different pole-zero locations as well as types (i.e., real or complex conjugate), in our simulation, different ARMA systems are considered. It is a common practice in modeling real data examples that zeros are placed between poles to char-

acterize the spectral notches. Identification problems when the zeros and poles are closely spaced have been considered in our simulation to demonstrate the efficacy of the proposed scheme in dealing with such a difficult situation. System pole locations also affect the decay in ACF. Generally, for natural signals, the ACF shows a moderate or very less decay which results more of its lags suitable for employing in the task of system identification. The proposed method has been tested for different systems having different pole locations. As the real and complex types of zeros or poles exhibit quite different behaviors, in our experiments various combinations of real and complex poles and zeros are considered to show the capability of the proposed algorithm in dealing with real life situations.

The effect of the data length on the derivation of the proposed model has also been addressed. It has been found that the autocorrelation function obtained by the proposed model is quite accurate even for a smaller data length. The estimation accuracy obtained by the proposed method is consistent throughout most of the data length. However, at a very low SNR, it is possible to achieve a significant improvement in the estimation accuracy at the expense of large data length by using the proposed ARMAC method.

In the proposed method, the RBLS optimization algorithm is employed for the estimation of AR and MA parameters. The computational complexity in terms of the arithmetic operations, namely, multiplication and addition, therefore, depends on the number of search points used in the RBLS optimization process. Consequently, the computational complexity of the proposed method is comparatively higher than that of the ACR or DYW method. The amount of computation for the ML method depends on the nonlinear optimization method, such as a Gauss-Newton type of algorithm. As mentioned earlier, in the proposed method, a significant reduction

in the frequency search space is achieved by restricting the search domain in the neighborhood of the initial frequency estimate. A further reduction in the search complexity has been achieved by employing a two-stage coarse and fine search scheme for the magnitude estimation, providing quite an accurate estimate of the parameters. The computational time required by the proposed method is found quite reasonable for practical applications where the objective is to achieve an accurate estimation of the ARMA system parameters at very low levels of SNR.

Extensive simulations have been carried out to demonstrate the performance of the proposed technique. It has been shown that the scheme outperforms the other existing methods that have been considered for comparison and is able to identify the ARMA parameters with sufficient accuracy and consistency in noisy environments ranging from very low to high levels of SNR. As a practical application of the proposed technique, the identification of human vocal-tract system in the presence of noise has been considered and shown to perform much better in comparison to other existing techniques in terms of the power spectral density of the resulting synthesized speech.

Some of the distinctive features of the proposed ARMAC method reinforced by the experimental results can be summarized as follows.

1. It is capable of handling the problem of ARMA system identification under heavy noisy conditions for both white noise and impulse-train input excitations. This feature makes the method readily applicable to speech signals, where either type of input excitations may occur.

2. The proposed noise-compensation scheme reduces the effect of additive noise in the MA parameter estimation, thus increasing the accuracy of the zero estimation.

3. In the proposed method, the input excitation power need not be assumed to be

known.

4. The proposed method estimates the ARMA parameters with guaranteed stability.

5. A byproduct of the proposed method is its availability of the estimated observation noise variance once the AR parameters have been obtained.

Chapter 3

Ramp Cepstrum Model Based System Identification

3.1 Introduction

In this chapter, cepstral domain approaches for the identification of a minimum-phase AR and ARMA systems in the presence of heavy noise are presented. First, two ramp-cepstrum models valid for both white noise and periodic impulse-train excitations are proposed for the one-sided autocorrelation function of AR and ARMA signals. The residue-based least-squares (RBLS) optimization technique, as described in Chapter 2, is then employed in conjunction with the ramp-cepstrum model to estimate the AR parameters of the AR or ARMA system from the noisy output observations, with a guaranteed system stability. The proposed ramp-cepstral model fitting combines the good features of both the correlation and cepstral domains, and thus provides a more accurate estimate of the parameters in a noisy environment [104]–[107]. In order to estimate the MA parameters of the ARMA system, two different approaches are developed. In the first approach, a one-step (OS) or direct scheme is proposed to estimate the MA parameters using the ramp-cepstrum model together with the RBLS algorithm in a similar manner as the AR parameter are

estimated. In the second approach, a two-stage (TS) scheme is proposed where a residual signal is first obtained by filtering the observed data via the estimated AR parameters. Then, the MA parameters are estimated from the residual signal by using the noise-compensated scheme proposed in Chapter 2. Extensive simulations are carried out on synthetic AR and ARMA systems of different orders in the presence of noise. Simulation results demonstrate quite a satisfactory identification performance even for an SNR of -5 dB, a level at which most of the existing methods fail to provide accurate estimation. To illustrate the suitability of the proposed technique in practical applications, the human vocal-tract system identification is also carried out using natural speech signals.

The rest of the chapter is organized as follows. In Section 3.2, the proposed AR system identification method based on the ramp-cesptrum model is described. First, a ramp-cepstrum model of a OSACF of an AR signal for the two types of input excitations is derived. AR parameter estimation scheme under noisy conditions is then described which utilizes the RBLS optimization algorithm in conjunction with the proposed ramp-cepstrum model. Simulation results on different synthetic AR systems and natural speech signal are presented in Section 3.3. In Section 3.4, the proposed ARMA system identification methods are presented. First, the ramp-cepstrum model of the OSACF of ARMA signal is developed and then two different approaches for the estimation of the ARMA system parameters are introduced. The estimation performance of the proposed ARMA system identification method is demonstrated in Section 3.5 through simulations for both synthetic and natural speech signals. Finally, in Section 3.6, some features of the proposed methods are summarized with some concluding remarks.

3.2 AR System Identification

3.2.1 Background

A causal stable linear time-invariant AR system can be characterized by

$$x(n) = -\sum_{i=1}^{P} a_i x(n-i) + u(n) \tag{3.1}$$

where $u(n)$ and $x(n)$ are, respectively, the excitation and response of the system, $\{a_i\}$ the AR parameters to be estimated, and P the system order assumed to be known. The transfer function of the AR(P) system can be written as

$$H(z) = \frac{1}{A(z)} = \frac{1}{1+\sum_{i=1}^{P} a_i z^{-i}} = \frac{1}{\prod_{i=1}^{P}(1-p_i z^{-1})} \tag{3.2}$$

where p_i represents the ith pole. The complex cepstrum (CC) of the impulse response $h(n)$ is defined as [8]

$$c_h(n) = F^{-1}\{\ln[F\{h(n)\}]\} = F^{-1}\{\ln[H(e^{j\omega})]\} \tag{3.3}$$

where F^{-1} denotes the inverse Fourier transform (FT). Since $h(n)$ is real and minimum phase, $c_h(n)$ is a sequence that is real and causal. From (3.2), $\ln[H(z)]$ can be expanded as

$$\ln[H(z)] = -\sum_{i=1}^{P} \ln\left(1-p_i z^{-1}\right) = \sum_{i=1}^{P}\sum_{n=1}^{\infty} \frac{p_i^n}{n} z^{-n} \tag{3.4}$$

where $|z| > |p_i|$. Thus, $c_h(n)$ can be expressed as

$$c_h(n) = \sum_{i=1}^{P} \frac{p_i^n}{n}, \quad n > 0 \tag{3.5}$$

Noting that $x(n) = h(n) * u(n)$, $c_x(n)$ can be written as

$$c_x(n) = c_h(n) + c_u(n) \tag{3.6}$$

Based on (3.6), a cepstrum domain AR SI method has been proposed in [73] for a noise-free environment. In the presence of additive noise $v(n)$, the observed signal $y(n)$ is given by

$$y(n) = x(n) + v(n) \tag{3.7}$$

where $v(n)$ is assumed to be zero mean stationary and inde-pendent of $u(n)$. The CC of $y(n)$ can then be expressed as

$$\begin{aligned} c_y(n) &= F^{-1}\{\ln[X(e^{j\omega})]\} + F^{-1}\left\{\ln\left[1 + \frac{V(e^{j\omega})}{X(e^{j\omega})}\right]\right\} \\ &= c_x(n) + c_w(n) \end{aligned} \tag{3.8}$$

where $c_w(n)$ arises from the presence of $v(n)$, and vanishes in its absence. The AR SI method faces a problem of obtaining an accurate estimate of $c_x(n)$ from $c_y(n)$, since the cepstrum decomposition techniques [71] are very sensitive to the noise level. In order to overcome the problem of observation noise and identify the AR system accurately under noisy condition, in what follows first, we propose a ramp cepstrum model and then based on the new model a residue based least-squares algorithm is developed to estimate the system parameters.

3.2.2 Proposed AR Ramp Cepstrum Model

(a) White Noise Excitation: In order to compute the cepstral coefficients, we propose to utilize the one-sided ACF (OSACF) given by

$$r_x(\tau) = \begin{cases} \rho_x(\tau), & \tau > 0 \\ 0.5\rho_x(\tau), & \tau = 0 \\ 0, & \tau < 0 \end{cases} \tag{3.9}$$

where $\rho_x(\tau)$ is the two-sided ACF of $x(n)$. Note that $r_x(\tau)$ is real and retains the pole-preserving property of $\rho_x(\tau)$, implying that all the information contained in $\rho_x(\tau)$ is maintained by $r_x(\tau)$. Moreover, $r_x(\tau)$ exhibits a higher noise immunity than $\rho_x(\tau)$

does [108]. Since, $\rho_x(\tau) = r_x(\tau) + r_x(-\tau)$, the FT of $\rho_x(\tau)$, i.e., the PSD of $x(n)$ can be expressed in terms of $R_x(e^{j\omega})$, the FT of $r_x(\tau)$, as

$$P_x(e^{j\omega}) = 2\Re e[R_x(e^{j\omega})] \tag{3.10}$$

and the CC corresponding to $P_x(e^{j\omega})$ can be written as

$$c_{P_x}(n) = F^{-1}\{\ln(\Re e[R_x(e^{j\omega})]) + \ln(2)\} \tag{3.11}$$

On the other hand, $P_x(e^{j\omega})$ is related to $H(e^{j\omega})$ as

$$P_x(e^{j\omega}) = \left|H(e^{j\omega})\right|^2 P_u(e^{j\omega}) \tag{3.12}$$

Thus, $c_{P_x}(n)$ given by (3.11) can be expressed as

$$c_{P_x}(n) = \begin{cases} c_h(n) + c_{P_u}(n), & n > 0 \\ 2c_h(0) + c_{P_u}(n), & n = 0 \\ c_h(-n) + c_{P_u}(n), & n < 0 \end{cases} \tag{3.13}$$

Note that in (3.11), the inverse FT of the term $\ln(2)$ vanishes for $n > 0$, and from (3.11) and (3.13) for $n > 0$ we can have

$$\mu_x(n) = F^{-1}\{\ln(\Re e[R_x(e^{j\omega})])\} = c_h(n) + c_{P_u}(n), n > 0 \tag{3.14}$$

For $n > 0$, $c_{P_x}(n)$ is now explicitly termed as $\mu_x(n)$ given by (3.14). In a fashion similar to (3.6), we have been able to establish a relation between the CC of the OSACF of $x(n)$ and $c_h(n)$. When $u(n)$ is a white Gaussian noise with $P_u(e^{j\omega}) = \sigma_u^2$, $c_{P_u}(n)$ vanishes except for $n = 0$, and then (3.14) reduces to

$$\mu_x(n) = c_h(n) = \sum_{i=1}^{P} \frac{p_i^n}{n}, n > 0 \tag{3.15}$$

It is seen that $\mu_x(n)$ decays rapidly with n, which makes it difficult to estimate the system poles from $\mu_x(n)$. In order to overcome this problem, we propose an easy-to-handle ramp-cepstrum (RC) for the OSACF which is defined as

$$\psi_x(n) = n\mu_x(n) = \sum_{i=1}^{P} p_i^n, n > 0 \tag{3.16}$$

Note that the system can have poles that are real or complex conjugate. Thus, in (3.16), each real pole or complex pole pair produces one exponentially decaying term, leading to

$$\psi_x(n) = \sum_{i=1}^{\gamma} \beta(\omega_i) r_i^n \cos(\omega_i n), n > 0 \tag{3.17}$$

where γ 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, and $\beta(\omega_i)$ is given by

$$\beta(\omega_i) = \begin{cases} 1, & \omega_i = 0 \text{ or } \omega_i = \pi \\ 2, & 0 < \omega_i < \pi \end{cases} \tag{3.18}$$

which is introduced to distinguish between real and complex poles. The model given by (3.17) is termed as the AR ramp-cepstrum (ARRC) model for the OSACF of $x(n)$, which will be used to form an objective function for the LS fitting. In this case, an estimate of $\rho_x(\tau)$ can be obtained, in general, as

$$\rho_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} x(n)x(n+|\tau|), 0 \leq |\tau| < N \tag{3.19}$$

where N is the data length [33]. This equation provides an accurate estimate of $\rho_x(\tau)$ when N is sufficiently large.

(b) Periodic Impulse-train Excitation: A periodic impulse-train excitation of length N $\{u_i(n), n \in (0 : N-1)\}$ with period T can be expressed as

$$u_i(n) = \sum_{k=0}^{\lambda-1} \delta(n - kT), \lambda = \left\lceil \frac{N}{T} \right\rceil \tag{3.20}$$

where $\lceil \zeta \rceil$ represents the lowest integer greater than or equal to ζ and λ is the total number of impulses within the excitation. Using (3.19), an estimate of the ACF of $u_i(n)$ is obtained as

$$\rho_{u_i}(\tau) = \sum_{k=0}^{\lambda-1} \left[\frac{\lambda - k}{N} \right] \delta(|\tau| - kT), 0 \leq |\tau| < N \tag{3.21}$$

It can be shown that (3.21) can be rewritten as

$$\rho_{u_i}(\tau) = \begin{cases} g(\tau/T), & |\tau| = 0, T, 2T, \ldots, (\lambda-1)T \\ 0, & \text{otherwise} \end{cases} \tag{3.22}$$

where

$$g(n) = \begin{cases} (\lambda - |n|)/N, & |n| \leq \lambda - 1 \\ 0, & \text{otherwise} \end{cases} \tag{3.23}$$

It is evident that $\rho_{u_i}(\tau)$ can be obtained by upsampling $g(\tau)$ with a factor T. The FT of $\rho_{u_i}(\tau)$ can be expressed as

$$P_{u_i}(e^{j\omega}) = G(e^{j\omega T}) \tag{3.24}$$

where $G(e^{j\omega})$ is the FT of $g(n)$ given by

$$G(e^{j\omega}) = \frac{1}{N}\left[\frac{\sin(\omega\lambda/2)}{\sin(\omega/2)}\right]^2 \tag{3.25}$$

Since, $P_{u_i}(e^{j\omega})$ and $ln[P_{u_i}(e^{j\omega})]$ are periodic with a period $2\pi/T$, the CC corresponding to $P_{u_i}(e^{j\omega})$ can be written as

$$c_{P_{u_i}}(n) = \begin{cases} c_g(n/T), & n = 0, T, 2T, \ldots, (\lambda-1)T \\ 0, & \text{otherwise} \end{cases} \tag{3.26}$$

where $c_g(m)$ is the CC of $g(m)$. Thus, it is clear from (3.14) and (3.26) that $c_{P_{u_i}}(n)$ contributes to $\mu_x(n)$ only at the origin or when $n \geq T$. Therefore, (3.14) can be reduced to

$$\mu_x(n) = c_h(n) = \sum_{i=1}^{P} \frac{p_i^n}{n}, 0 < n < T \tag{3.27}$$

From (3.15) and (3.27), it can be observed that the ARRC model derived for the white noise excitation is also valid for the case of impulse-train excitation when $0 < n < T$.

3.2.3 Ramp Cepstral Fitting: Residue Based Least-Squares Minimization

In the presence of noise $v(n)$, the ACF of the noisy observation $y(n)$ can be expressed as

$$\rho_y(\tau) = \rho_x(\tau) + \rho_v(\tau) + \rho_{xv}(\tau) + \rho_{vx}(\tau) \tag{3.28}$$

Thus, we can see that the component due to the noise, $\rho_n(\tau) = \rho_v(\tau) + \rho_{xv}(\tau) + \rho_{vx}(\tau)$, corrupts $\rho_x(\tau)$. This effect cannot be neglected, especially when the SNR is very low. Hence, the conventional correlation based methods using $\rho_y(\tau)$, cannot provide a good estimation performance. As in the case of the cepstrum in the signal domain (see (3.8)), we can have the cepstrum representation of (3.28) in the correlation domain as

$$\begin{aligned} c_{P_y}(n) &= F^{-1}\{\ln[P_x(e^{j\omega})]\} + F^{-1}\left\{\ln\left[1 + \frac{P_n(e^{j\omega})}{P_x(e^{j\omega})}\right]\right\} \\ &= c_{P_x}(n) + c_{P_w}(n) \end{aligned} \tag{3.29}$$

where $c_{P_w}(n)$ is the CC introduced due to the noise. By using the OSACF $r_y(\tau)$ of $y(n)$, defined in a manner similar to (3.9), (3.29) can be modified as

$$\mu_y(n) = F^{-1}\{\ln(\Re e[R_y(e^{j\omega})])\} = \mu_x(n) + c_{P_w}(n), n > 0 \tag{3.30}$$

Therefore, the RC of $r_y(\tau)$ can be expressed as

$$\psi_y(n) = \psi_x(n) + \psi_w(n), n > 0 \tag{3.31}$$

Here, $\psi_w(n)$ is the error introduced due to the noise. When the additive noise is white Gaussian, the zero lag of the noisy ACF is most severely corrupted. One way of reducing the effect of noise is to replace the actual value $\rho_y(0)$ by a smaller value while computing $\psi_y(n)$ from $r_y(\tau)$. Note that $\rho_y(0) > |\rho_y(\tau)|$ for $\tau \neq 0$. In order to

reduce the effect of noise, we replace $\rho_y(0)$ by $\eta\rho_y(0)$ with $\{|\rho_y(1)|/\rho_y(0)\} \leq \eta < 1$. The process can efficiently suppress the level of $\psi_w(n)$ while leaving the shape of $\psi_y(n)$ similar to that of $\psi_x(n)$.

The AR parameters can be determined by estimating the parameters $\{r_i\}$ and $\{\omega_i\}$ of the ARRC model given in (3.17). Each of the γ component functions in (3.17) is estimated sequentially from M_c instances of $\psi_y(n)$, where $M_c < T$ for the impulse-train excitation. The objective function is defined as the total squared error between the $(l-1)$th residual function $\Re_{l-1}(n)$ and the lth component of the model, that is

$$J_l = \sum_{n=1}^{M_c} |\Re_{l-1}(n) - \beta(\omega_l) r_l^n \cos(\omega_l n)|^2 , l = 1, \ldots, \gamma \tag{3.32}$$

where the residual function is updated as follows

$$\begin{aligned} \Re_0(n) &= \psi_y(n) \\ \Re_l(n) &= \Re_{l-1}(n) - \beta(\omega_l) r_l^n \cos(\omega_l n),\ l = 1, \ldots, \gamma - 1 \end{aligned} \tag{3.33}$$

We would like to find the optimal solution for $\{r_l\}$ and $\{\omega_l\}$ by a search algorithm. For each set of the chosen values of $\{r_l\}$ and $\{\omega_l\}$, the values corresponding to the global minimum of J_l are selected as the estimate of the desired poles. Proceeding this way, the AR parameters can be determined using (3.2) once all the P 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. In order to reduce the computational burden, a two-step search algorithm can be used. In the first step, only a coarse-search based on the fast FT (FFT) is employed to find out the initial estimate of $\{\omega_l\}$ and $\{r_l\}$, and in the second step, a fine-search is carried out around the initial estimate with a higher resolution to obtain an accurate estimate.

3.3 Simulation Results on AR System Identification

Computer simulations are carried out on several synthetic AR and ARMA systems as well as on the natural speech signals. In this section, first, some simulation results on AR system identification are presented and the estimation performance of the proposed method is compared with that of some of the existing methods, namely, improved least-squares with direct implementation (ILSD) structure giving a faster convergence [62], signal/sub-space Yule-Walker (SSYW) [58], modified least-squares YW (MLSYW) [33], and the spectral all-pole estimation (SAPE) [109] considering AR systems of different orders and pole locations. Then we present some simulation results to compare the performance of the proposed ARMA system identification method with that of the order-selective Durbin's (OSD) method [110] and the ACR method [45]. In the OSD and ACR methods, AR parameters are estimated using the LSMYW method [33]. For the estimation of MA part, the ACR method employs the so-called ARMA-cepstrum recursion while the OSD method pursues the intermediate AR parameter estimation using the LSMYW equations.

3.3.1 White Noise Excitation

A noisy signal is generated according to (3.1) and (3.7) with $N = 4,000$ and $\sigma_u^2 = 1$, where the noise variance σ_v^2 is appropriately determined according to a specified level of SNR defined as

$$\text{SNR} = 10\log_{10}(\sum_{n=0}^{N-1} x(n)^2 / \sum_{n=0}^{N-1} v(n)^2) \quad \text{dB} \tag{3.34}$$

An initial estimate of ω_l is obtained from the location of the peaks of the smoothed FFT of the OSACF of $y(n)$. In our simulation, the search range for ω_l is chosen

Table 3.1: Performance comparison at SNR = −5 dB for AR(3) system

True	Proposed	ILSD	SSYW	MLSYW	SAPE
a_1=	−2.3149	−1.2619	−0.6259	−1.0445	−1.9181
−2.2990	±0.0841	±0.6346	±0.3086	±0.0923	±0.1432
	±0.0856	±1.2158	±1.6859	±1.2579	±0.4018
a_2=	2.1493	0.6324	−0.2562	0.4682	1.1196
2.1262	±0.1318	±1.0318	±0.0611	±0.1136	±0.1666
	±0.1338	±1.8155	±2.3832	±1.6619	±0.9719
a_3=	−0.7676	−0.0638	0.3425	−0.4232	−0.2217
−0.7604	±0.0816	±0.5959	±0.0613	±0.5910	±0.0566
	±0.0819	±0.9167	±1.1715	±0.6804	±0.6804
ASSE	−19.73	2.72	4.21	3.36	−2.60

as $\pm 0.05\pi$ of the initial estimate and that for r_l as $0.5 \leq r_l \leq 0.99$, with search resolutions of $\Delta\omega = 0.01$ and $\Delta r = 0.01$. We take $M_c = 20P$ and $\eta = |\rho_y(1)|/\rho_y(0)$. Each experiment contains $N_T = 100$ independent trials, where the SNR varies from −10 dB to 15 dB. The criteria used for the performance measurement are (1) the mean of estimated parameters, (2) the standard deviation from the mean (SDM), (3) the SD from the true value (SDT), and (4) the average sum-squared error (ASSE) given by

$$\text{ASSE} = \frac{1}{N_T P} \sum_{m=1}^{N_T} \sum_{k=1}^{P} [\hat{a}_k(m) - a_k]^2 \tag{3.35}$$

where $\hat{a}_k(m)$ represents the estimated parameter at the mth trial.

The estimation results for the AR(3) and AR(4) systems at SNR = −5 dB are shown, respectively, in Tables 3.1 and 3.2, where each row, following the true value of a parameter, lists the mean, SDM, and SDT of the estimated values for different methods. The last row provides the ASSE in dB. It is seen from these tables that the proposed method exhibits superior performance with respect to all the four performance indices. Clearly, the small values of SDM and SDT obtained from the proposed technique indicate a high degree of estimation consistency and accuracy. The effect of SNR on the estimation error is shown in Fig. 3.1 for the AR(3) sys-

Table 3.2: Performance comparison at SNR = −5 dB for AR(4) system

True	Proposed	ILSD	SSYW	MLSYW	SAPE
a_1=	0.5937	0.5656	0.4967	0.0393	0.0452
0.6349	±0.0598	±0.1787	±0.3969	±0.0745	±0.0336
	±0.0687	±0.1823	±0.4097	±0.6783	±0.5907
a_2=	−0.0631	−0.0671	−0.1043	0.0386	−0.1762
−0.0416	±0.0425	±0.0434	±0.2746	±0.0528	±0.0139
	±0.0513	±0.0491	±0.2816	±0.0960	±0.1354
a_3=	−0.8105	−0.7734	−0.6806	−0.8482	−0.0419
−0.8100	±0.0503	±0.0802	±0.2675	±0.0310	±0.0329
	±0.0503	±0.0819	±0.2971	±0.0492	±0.7688
a_4=	−0.6904	−0.6579	−0.5523	−0.0717	−0.2863
−0.7219	±0.0498	±0.1922	±0.2941	±0.0383	±0.0105
	±0.0871	±0.1972	±0.3058	±0.6513	±0.4357
ASSE	−22.99	−16.47	− 9.15	− 6.49	− 4.42

tem used in Table 3.1. It is observed from this figure that in comparison to the proposed method, ILSD and SSYW methods exhibit a similar estimation performance for SNR over 10 dB. Note that the SAPE method is for the identification of AR systems in the noise-free case and thus not expected to perform well under a low level of SNR. The proposed method performs significantly better even for the low levels of SNR. In Fig. 3.2, the average estimated poles obtained from the proposed method at SNR = −5 dB along with their true locations are plotted for an AR(4) system with $a_k = \{1, -2.595, 3.339, -2.2, 0.731\}$ and for an AR(6) system with $a_k = \{1, -2.0825, 2.267, -2.1997, 1.8563, -0.811, 0.8563\}$. It is seen from these plots that the proposed method provides a very high estimation accuracy. It may be pointed out that the estimation results of the other methods for these systems at SNR = −5 dB are not as accurate.

3.3.2 Impulse-Train Excitation

An impulse-train is generated using (3.20) with $N = 4,000$. We choose $M_c = min(T - 1, 20P)$. Fig. 3.3 shows the ASSE-SNR plots for the same AR(3) system as considered

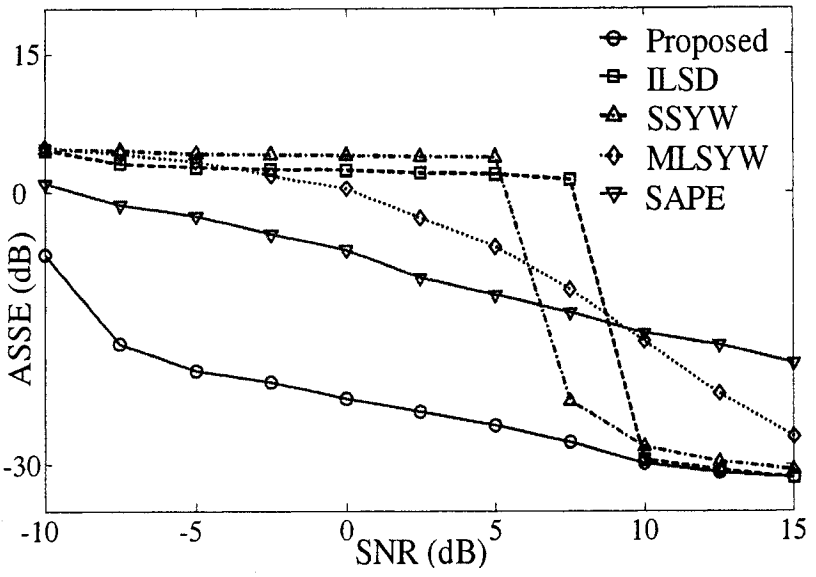


Figure 3.1: Effect of noise level on the ASSE for a white noise excited system.

above with $T = 120$. As seen, the proposed method provides a significantly better performance even at SNR $= -5$ dB, whereas the performance of other methods deteriorates below SNR $= 7.5$ dB. Similar results are observed for other systems considered for the white noise excitation.

3.3.3 Application for Vocal-tract System Identification

Some English natural phonemes from the TIMIT standard database have been tested with a view to identify the vocal-tract (VT) system in the presence of additive noise. With the estimated VT AR parameters, the pitch-period (for voiced speech), and the AR filter gain, a speech phoneme can be synthesized [7]. It is to be mentioned that the synthesized sounds obtained by different methods were also played back in order to test the subjective quality. Fig. 3.4(a) shows the PSD of the synthesized speech obtained by using different methods along with that of the noise-free speech,

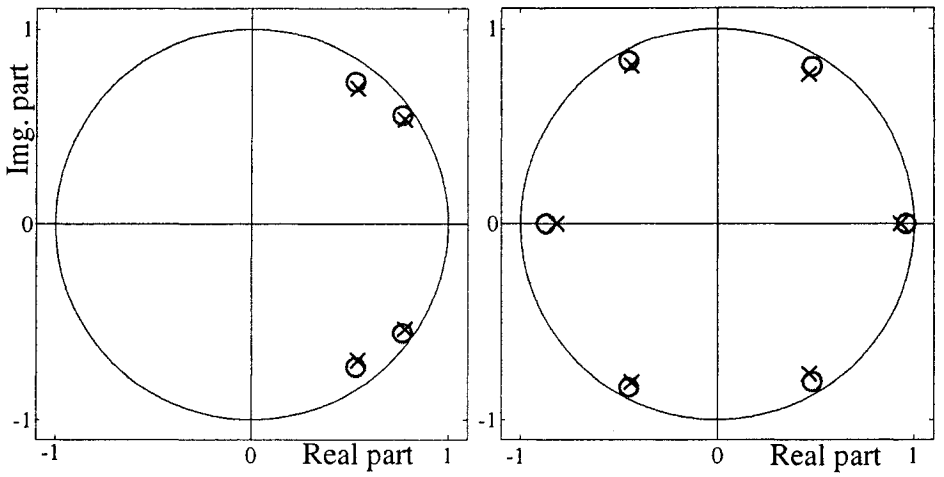


Figure 3.2: True and estimated poles at SNR = −5 dB for (a) AR(4) system, and (b) AR(6) system. (o : true poles, × : estimated poles).

and that obtained from one of the 20 noisy signals under a white Gaussian noise of SNR = −5 dB. In this case, an AR(10) model is used for a female sound /*a*/ of the word '*Rob*'. According to the general behavior of the VT parameter, r_l is searched in the range $[0.8, 0.99]$ [93] and the search range for ω_l can be narrowed down based on the knowledge of the locations of the VT resonances [7], [93]. Quality of the synthesized sounds obtained by the proposed method under such a noisy condition were found far superior than that obtained by the other methods. Fig. 3.4(b) gives the corresponding PSD results for the case of a multi-talker babble noise (multiple background competing speakers) with SNR = −5 dB. It is observed that the proposed method provides satisfactory performance for both the white noise and babble noise under very low levels of SNR.

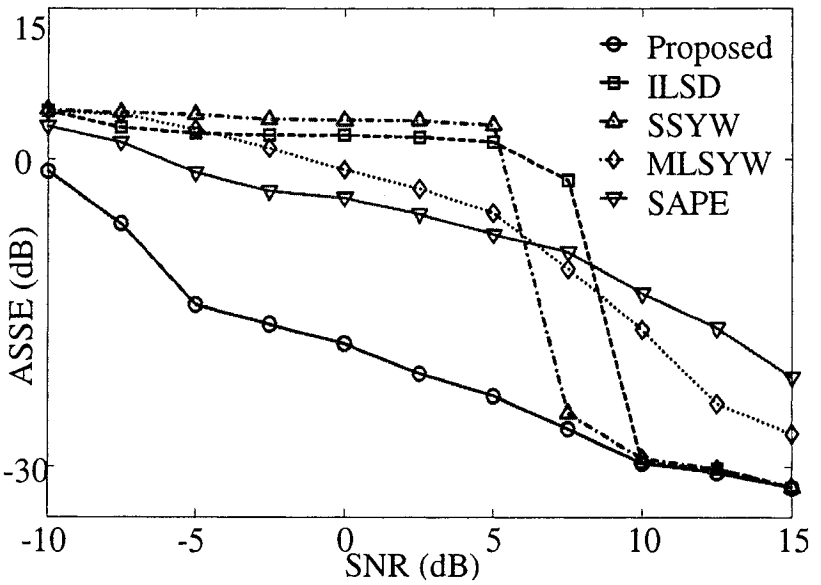


Figure 3.3: Effect of noise level on the ASSE for an impulse-train excited system.

3.4 ARMA System Identification

In this section, we are going to develop the ramp-cepstrum model corresponding to the ARMA system. For the purpose of development, we follow a similar way which is adopted to obtain the ARRC model from the OSACF keeping in mind some obvious changes due to the presence of zeros apart from the poles. An ARMA system identification scheme is then developed based on the model-fitting approach.

3.4.1 Background

As described in Chapter 2, a causal stable and LTI ARMA (P, Q) system is characterized by

$$\sum_{i=0}^{P} a_i x(n-i) = \sum_{j=0}^{Q} b_j u(n-j) \tag{3.36}$$

where $u(n)$ and $x(n)$ are, respectively, the excitation and the response of the system, a_i and b_j the corresponding AR and MA parameters with $a_0 = 1$ and $b_0 = 1$, and

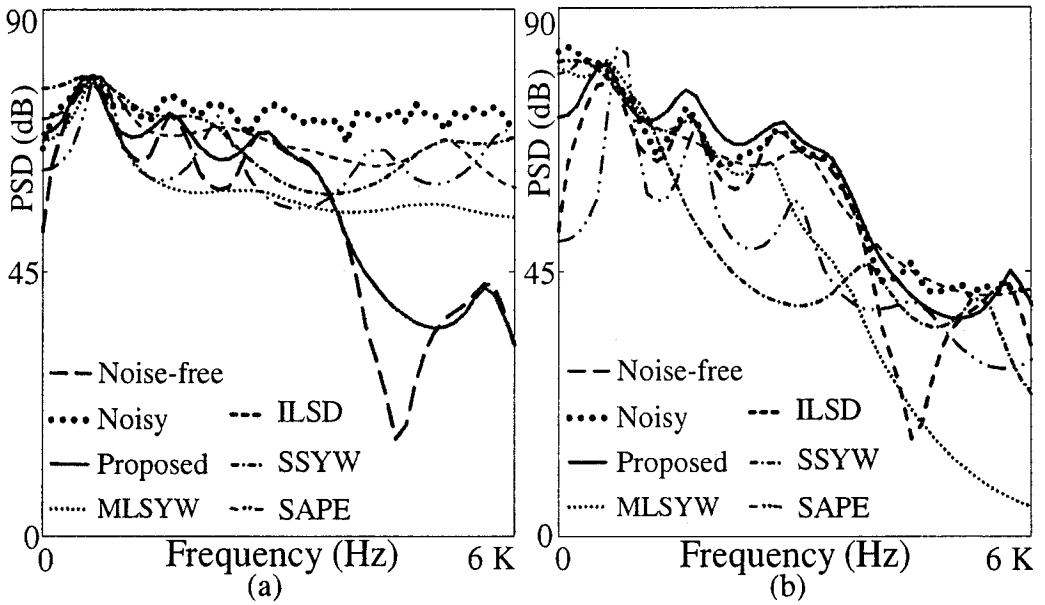


Figure 3.4: PSD obtained for a natural speech phoneme /a/ at SNR = −5 dB for the case of (a) white noise, and (b) multi-talker babble noise.

P and Q are the orders of the ARMA model, which are assumed to be knownThe corresponding system transfer function is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{\prod_{j=1}^{Q}(1 - z_j z^{-1})}{\prod_{k=1}^{P}(1 - p_k z^{-1})} = \sum_{k=1}^{P} \frac{\eta_k}{1 - p_k z^{-1}} \tag{3.37}$$

where $A(z) = \sum_{k=1}^{P} a_k z^{-k}$ and $B(z) = 1 + \sum_{j=1}^{Q} b_j z^{-j}$ are, respectively, the AR and MA polynomial, p_k's and z_j's denote, respectively, the poles and the zeros of the ARMA system. It is assumed that all poles and zeros are of the first-order and the the ARMA(P, Q) process is minimum phase and stationary. From (3.37), $\ln[H(z)]$

can be expanded as

$$\begin{aligned}\ln[H(z)] &= -\sum_{i=1}^{P}\ln\left(1-p_i z^{-1}\right)+\sum_{j=1}^{Q}\ln\left(1-z_j z^{-1}\right)\\ &= \sum_{i=1}^{P}\sum_{n=1}^{\infty}\frac{p_i^n}{n}z^{-n}-\sum_{j=1}^{Q}\sum_{n=1}^{\infty}\frac{z_j^n}{n}z^{-n}\end{aligned} \tag{3.38}$$

where $|z| > |p_i|$. Thus, the complex cepstrum (CC) of $h(n)$ $c_h(n)$ can be expressed as

$$c_h(n) = \sum_{i=1}^{P}\frac{p_i^n}{n}-\sum_{j=1}^{Q}\frac{z_j^n}{n},\ n>0 \tag{3.39}$$

As mentioned for the case of AR parameter estimation under noise-free condition, based on (3.6), an AR SI method was proposed in [73]. However, in the case of ARMA SI, $c_h(n)$ contains terms with system zeros along with poles, making the identification problem difficult. As an alternative, some recursive cepstral domain methods are proposed in [11], [45], where a priori estimate of parameters of one part of the ARMA system (either AR or MA) is required in order to estimate the other part. Hence, it is a challenging issue to extract both AR and MA parameters of the ARMA system from cepstral coefficients. The problem becomes more difficult in the presence of noise, as explained after (3.8) that the cepstrum decomposition techniques [71] are very sensitive to the noise level. With a view to estimate the system parameters under noisy condition, in what follows we propose a ramp cepstrum model and based on which an ARMA system identification method will be developed.

3.4.2 Proposed ARMA Ramp Cepstrum Model

In order to develop an ARMA cepstrum model of the one-sided ACF (OSACF) of the observed data, a similar approach as presented in the previous section is adopted. It is to be mentioned that the basic difference in the case of ARMA system with respect

to that involved in the AR system is the introduction of the terms corresponding to the system zeros as shown in (3.5) and (3.39).

For the ARMA signal with a white Gaussian noise $u(n)$ excitation, $\mu_x(n)$ in (3.15) can be obtained as

$$\mu_x(n) = c_h(n) = \sum_{i=1}^{P} \frac{p_i^n}{n} - \sum_{j=1}^{Q} \frac{z_j^n}{n}, n > 0 \tag{3.40}$$

For a periodic impulse-train excitation $\{u_i(n)\}_{n=0}^{N-1}$ with period T, as derived above for the AR system, it can be shown that $c_{P_{u_i}}(n)$ contributes to $\mu_x(n)$ only at the origin or when $n \geq T$. Thus, $\mu_x(n)$ can be written as

$$\mu_x(n) = c_h(n) = \sum_{i=1}^{P} \frac{p_i^n}{n} - \sum_{j=1}^{Q} \frac{z_j^n}{n}, 0 < n < T \tag{3.41}$$

It can be observed from (3.40) and (3.41) that the expressions for $\mu_x(n)$ differ only in terms of the range of n. In what follows, although only $n > 0$ is written, it should be considered as $0 < n < T$ for the case of impulse-train excitation. Corresponding ramp-cepstrum can be written as

$$\psi_x(n) = n\mu_x(n) = \sum_{i=1}^{P} p_i^n - \sum_{j=1}^{Q} z_j^n, n > 0 \tag{3.42}$$

For a real-valued $x(n)$, complex poles (zeros) will always appear as conjugate pairs. In (3.42), the complex pole (zero) pairs and real poles (zeros) will each contribute one decaying exponential, which can be written as a decaying cosine function, yielding

$$\psi_x(n) = \sum_{i=1}^{\gamma_P} \beta(\omega_{p_i}) r_{p_i}^n \cos(\omega_{p_i} n) - \sum_{j=1}^{\gamma_Q} \beta(\omega_{z_j}) r_{z_j}^n \cos(\omega_{z_j} n), n > 0 \tag{3.43}$$

where $\gamma_P(\gamma_Q)$ represents the number of real poles (zeros) plus the number of complex conjugate pole (zero) pairs, r_{p_i} and ω_{p_i} are, respectively, the magnitude and the argument of the ith pole p_i, and r_{z_j} and r_{z_j} and ω_{z_j} are, respectively, the magnitude

and the argument of the jth zero z_j. In (3.43), $\beta(\omega)$ is introduced to distinguish the real and complex zeros and poles, namely, $\beta(\omega) = 1$ if $\omega = 0$ or π, otherwise $\beta(\omega) = 2$. As a result, (3.43) can be expressed as

$$\psi_x(n) = \sum_{k=1}^{\gamma'=\gamma_P+\gamma_Q} \beta(\omega_k) r_k{}^n \cos(\omega_k n),\ n > 0 \tag{3.44}$$

where $\beta(\omega_k)$ can be written as

$$\beta(\omega_k) = \begin{cases} (-1)^m, & \omega_k = 0 \text{ or } \omega_k = \pi \\ (-1)^m 2, & 0 < \omega_k < \pi \end{cases} \tag{3.45}$$

with m given by

$$m = \begin{cases} 0, & k \leq \gamma_P \\ 1, & k > \gamma_P \end{cases} \tag{3.46}$$

The model given by (3.44) is termed as the ARMA ramp-cepstrum (ARMARC) model for the OSACF of $x(n)$ which will be used to form an objective function for the LS fitting. Note that in case of impulse-train excitation as we mentioned in (3.44), $0 < n < T$ has to be considered instead of $n > 0$.

It is clear from (3.44) that the ARMA system parameters can be estimated by conducting the estimation of the ARMARC model parameters $\{r_k\}$ and $\{\omega_k\}$. Each of the γ' component functions in (3.44) is estimated sequentially from M_c nonzero instances of $\psi_y(n)$, where $M_c < T$ for the impulse-train excitation. The objective function can be formulated in a similar fashion as it is done for the AR system identification. Therefore, the total squared error between the $(l-1)$th residual function $\Re_{l-1}(n)$ and the lth component of the model is given by

$$J_l = \sum_{n=1}^{M_c} \left| \Re_{l-1}(n) - \beta(\omega_l) r_l^n \cos(\omega_l n) \right|^2, l = 1, \ldots, \gamma' \tag{3.47}$$

where the residual function is updated as follows

$$\begin{aligned} \Re_0(n) &= \psi_y(n) \\ \Re_l(n) &= \Re_{l-1}(n) - \beta(\omega_l) r_l^n \cos(\omega_l n),\ l = 1, \ldots, \gamma' - 1 \end{aligned} \tag{3.48}$$

We would like to find the optimal solution for $\{r_l\}$ and $\{\omega_l\}$ by a search algorithm similar to one that is described for the case of AR systems. For each set of the chosen values of $\{r_l\}$ and $\{\omega_l\}$, the values corresponding to the global minimum of J_l are selected as the estimate of the desired poles. Proceeding this way, the AR parameters can be determined using (3.2) once all the P 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 system. The major difference that has to be carefully noticed is that, as seen from (3.44), when $l \leq \gamma_P$, the estimated values of values of $\{r_l\}$ and $\{\omega_l\}$ correspond to a pole, otherwise when $l > \gamma_P$, a zero can be estimated. The two-step search algorithm described before can be also employed to reduce the computational burden. Since, ACF is a pole-preserving function and the effect of additive noise is more pronounced on system zeros rather than system poles, a better estimate of pole-locations is expected. Hence, the search operation can be restricted around the initial estimates instead of the entire domain of ω_l. On the other hand, search for zeros can be performed using a coarse search to obtain an initial estimate followed by finer search around the initial estimate. The stability of the estimated ARMA model can easily be guaranteed by restricting the search range of r_l.

It is to be mentioned that the residual ramp-cepstrum at the beginning of the estimation of zeros, namely,

$$\Re_{\gamma_P}(n) = [\psi_x(n) - \sum_{l=1}^{\gamma_P} \beta(\omega_l) r_l^n \cos \omega_l n] + \psi_n(n) \tag{3.49}$$

contains low energy in the signal part as a bulk of energy has already been removed by the component functions associated with the poles. This might cause error in the estimation of MA parameters at a very low SNR as observed in (3.49). As

an alternate of the above one-step (OS) method, a two-step (TS) MA parameter estimation algorithm proposed in Chapter 2 can easily be employed once the AR parameters are obtained using the ARMA ramp-cepstral least-square minimization. In this case, first, a residual signal $f_y(n)$ is obtained by filtering the noisy observed signal $y(n)$ via the estimated AR parameters. Then the zeros are computed from the IACF $\hat{\phi}_{f_x}(\tau)$ corresponding to the noise-compensated ACF $\hat{r}_{f_x}(\tau)$ of the residual signal using the RBLS algorithm.

3.5 Simulation Results on ARMA System Identification

Simulations are carried out for the identification of ARMA systems under noisy conditions, and results along with some comparative analysis are investigated in this section. Different ARMA systems with various pole-zero locations within the unit circle are considered. Next, the performance in terms of the accuracy and consistency of the estimated parameters of the proposed method is obtained and compared with that of the ARMA cepstrum recursion (ACR) method [45], and an order-selective Durbin's (OSD) method in which the Durbin's scheme for the MA parameter estimation is combined with the least-squares modified Yule-Walker (LSMYW) equations for the AR parameter estimation [33]. In the ACR method, AR parameters are also estimated using the LSMYW equations. For the estimation of MA part, the ACR method employs the so-called ARMA-cepstrum recursion while the OSD method pursues the intermediate AR parameter estimation using the LSMYW equations [97].

An ARMA signal is generated according to (3.7) and (3.36) with $N = 4,000$ and $\sigma_u^2 = 1$, where the noise variance σ_v^2 is appropriately determined according to a specified level of SNR defined in (3.34). The search range ω_l and r_l are kept same

Table 3.3: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited ARMA(4, 3) system at SNR = −5 dB

Methods	Estimated parameters							ASSE
	a_1	a_2	a_3	a_4	b_1	b_2	b_3	(dB)
Proposed	−2.5771	3.33859	−2.2829	0.7945	−1.9851	1.8624	−0.6758	−15.15
(TS)	(±0.1269)	(±0.2536)	(±0.2246)	(±0.0853)	(±0.0928)	(±0.1027)	(±0.0759)	
	(±0.1281)	(±0.2579)	(±0.2470)	(±0.1126)	(±0.01037)	(±0.1051)	(±0.0782)	
Proposed	−2.5771	3.3859	−2.2829	0.7945	−1.9748	1.7928	−0.5986	−12.33
(OS)	(±0.1269)	(±0.2536)	(±0.2246)	(±0.0853)	(±0.2316)	(±0.2841)	(±0.0959)	
	(±0.1281)	(±0.2579)	(±0.2470)	(±0.1126)	(±0.2689)	(±0.211)	(±0.0985)	
OSD	−0.6300	0.0914	0.1891	−0.1116	−0.4909	0.1258	0.0701	5.88
	(±0.1185)	(±0.1580)	(±0.1384)	(±0.0922)	(±0.5152)	(±0.4191)	(±0.2625)	
	(±1.9685)	(±3.2514)	(±2.3931)	(±0.6263)	(±1.6821)	(±1.7683)	(±0.7646)	
ACR	−0.6300	0.0914	0.1891	−0.1116	−0.5234	0.0068	0.1076	5.87
	(±0.1185)	(±0.1580)	(±0.1384)	(±0.0922)	(±0.1188)	(±0.1462)	(±0.1247)	
	(±1.9685)	(±3.2514)	(±2.3931)	(±0.6263)	(±1.5732)	(±1.8428)	(±0.7659)	
True	−2.5950	3.3390	−2.2000	0.7310	−2.0922	1.8438	−0.6480	

as that used for the AR system identification. Each experiment contains $N_T = 100$ independent realizations and we compute the estimation mean, SDM, SDT, and ASSE as defined before. An initial estimate of ω_l is obtained from the location of the peaks of the smoothed FFT of the OSACF of $y(n)$.

Table 3.3 shows the estimation results for an ARMA(4, 3) system at SNR −5 dB. In this table, the last row gives the true parameter values. The estimated values of the corresponding parameters obtained from the proposed one-step (OS) and two-step (TS) methods and two other methods are given in the preceding four rows. The corresponding values for the SDM and SDT are shown within the parentheses below the each estimated parameter value. It is found that both TS and OS methods are capable of providing quite satisfactory parameter estimation performance in comparison to the other methods which at such a low level of SNR completely fail to identify the system. The small values of SDM and SDT obtained by the proposed techniques indicate respectively a high estimation consistency and accuracy. It is to be noted that in the two-step ARMA identification technique, if the estimation error of AR

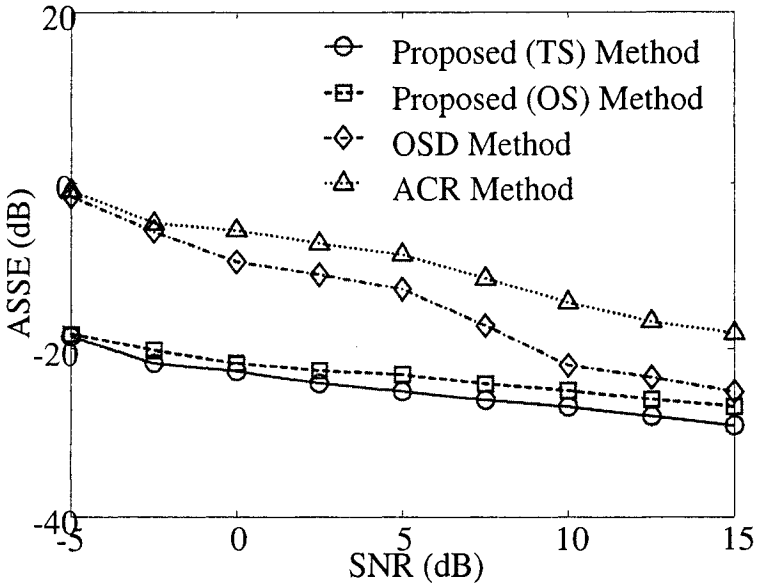


Figure 3.5: Effect of SNR on the estimation accuracy for the white noise excited ARMA(3, 2) system.

parameters in the first step is large, it will propagate in the next step and deteriorates MA estimation accuracy. This is one of the reasons behind the complete failure of the other two methods at very low SNRs. It is observed from our experimentation that the performance of the TS method is slightly superior than that of the OS method which supports the explanation given before.

Fig. 3.5 presents the variation of ASSE values with respect to the level of SNR for all four methods for an ARMA(3, 2) system with true parameters $a_k = \{1, -2.5712, 2.5218, -0.9460\}$ and $b_j = \{1, -1.6909, 0.81\}$. The OSD and ACR methods give a comparable estimation accuracy only at high levels of SNR. However, they show a poor performance when the SNR is low. The proposed TS and OS methods exhibit excellent performance even at a low level of SNR of 0 dB or even lower than that.

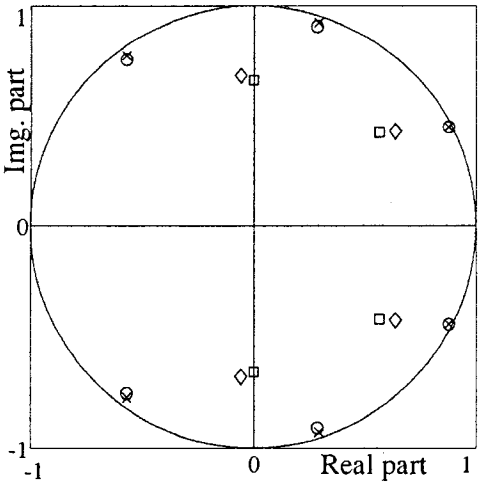


Figure 3.6: Estimated pole-zero plot of ARMA(6, 4) system obtained by the TS method at SNR = −5 dB; (Poles: ×: true, O: proposed; Zeros: □: true, ◇ :proposed).

In Figs. 3.6 and 3.7, the average estimated poles and zeros of an ARMA(6, 4) system obtained by the proposed TS and OS methods at SNR = −5 dB are shown respectively. The ARMA(6, 4) system parameters are

$$a_k = \{1, -1.2174, 1.2225, -1.0537, 1.1167, -1.0253, 0.85\}$$

$$b_j = \{1, -1.1354, 0.9399, -0.4974, 0.2198\}.$$

In this figure the true poles and zeros are also shown for the purpose of comparison. From this figure it can be observed that the proposed methods are also able estimate the poles and zeros quite accurately at a low level of SNR.

Identification performance of the proposed methods are also tested for the impulse-train excited systems under noisy conditions. A periodic impulse train with a known value of T is generated in a same manner as done for the AR part. In Fig. 3.8, the ASSE plots as a function of levels of SNR obtained by different methods for the same ARMA(3, 2) system as considered in Fig. 3.5 under the impulse-train excitation are shown. Here T is chosen as 68 and number of ramp-cepstral coefficients

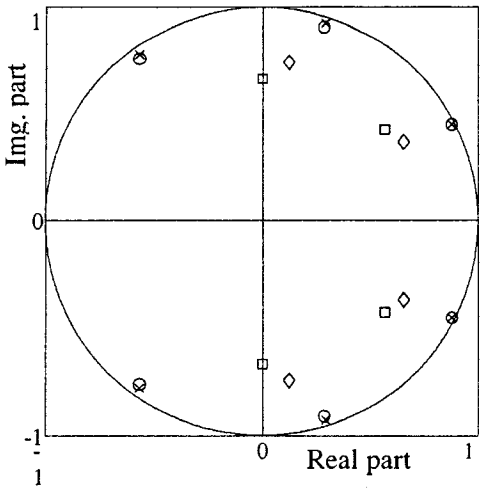


Figure 3.7: Estimated pole-zero plot of ARMA(6, 4) system obtained by the OS method at SNR = −5 dB; (Poles: ×: true, O: proposed; Zeros: □: true, ◇ :proposed).

$M_c = min(T/2, 10P)$. As seen, the proposed methods provide a significantly superior performance at a very low SNR in comparison to the other two methods.

As an application of the proposed ramp-cepstrum based ARMA system identification method, the estimation of a vocal tract system parameters is performed using natural speech signal in a similar manner as it was presented in Chapter 2. Some English nasal sounds (voiced phonemes) from the TIMIT standard database are used for testing. No pre-filtering is performed in order to observe the accuracy of the pole-zero estimation over the entire range of frequency. In order to verify the estimation accuracy, the PSD of the synthesized speech is compared with that of the noise-free natural speech. The synthesized sounds obtained by different methods were also played back and the quality of the synthesized sounds obtained by the proposed method under such a noisy condition were found far superior than that obtained by the other methods. Fig. 3.9 shows the PSD of the synthesized speech obtained by using different methods along with that of the noise-free speech, and that obtained

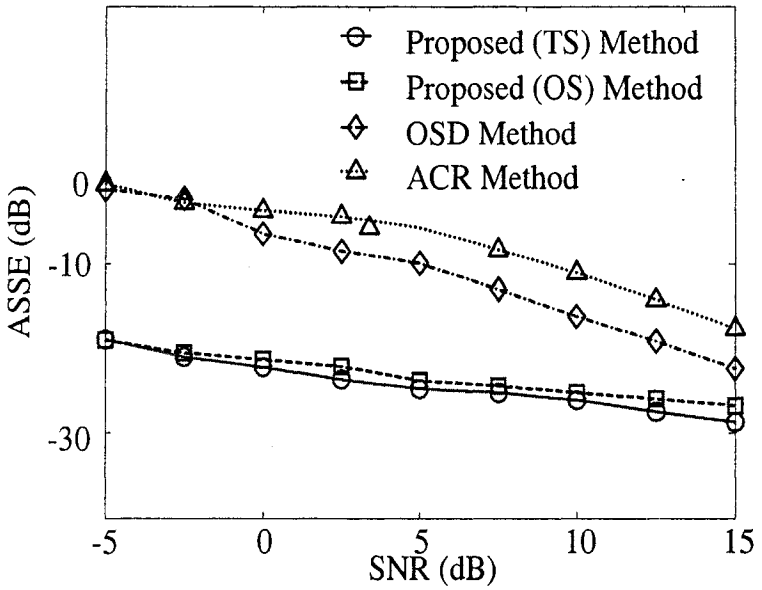


Figure 3.8: Effect of SNR on the estimation accuracy for the impulse-train excited ARMA(3, 2) system.

from one of the 20 noisy signals under a white Gaussian noise of SNR $= -5$ dB. Here we consider an ARMA$(12, 6)$ model, for a naturally spoken nasal sound $/m/$ of the word "him", uttered by a female speaker. The ARMARC samples are considered up to T/2 where the excitation period or pitch (T) can be estimated using available techniques. It is seen from Fig. 3.9 even at an SNR of $= -5$ dB, where the PSD of noise-free speech is completely covered by that of noisy speech, the proposed method exhibits a satisfactory performance in comparison to other methods.

3.6 Conclusion

In this chapter, new schemes for the identification of AR and ARMA systems from noise-corrupted output observations have been presented. A simple yet accurate AR ramp cepstrum (ARRC) model of the OSACF of an AR signal has been developed

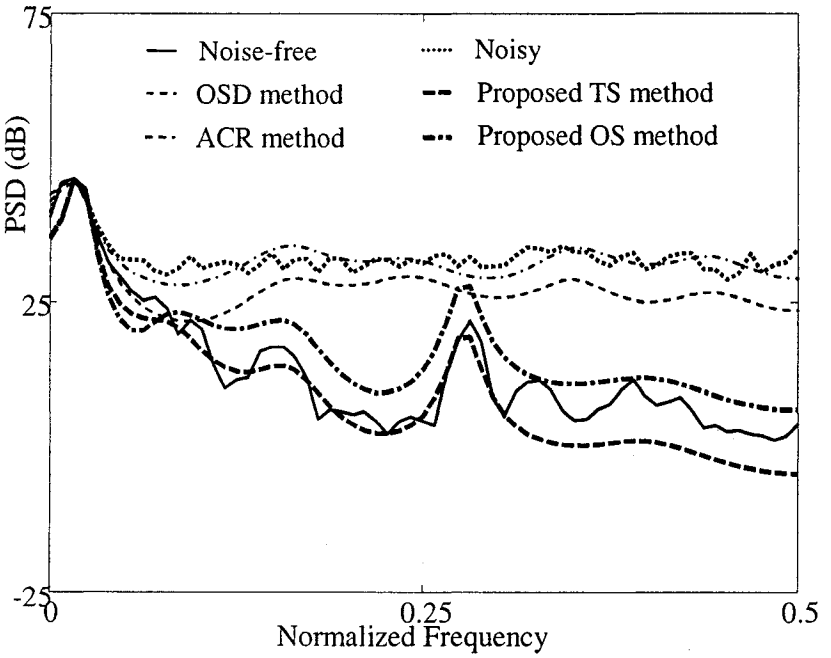


Figure 3.9: PSD obtained by using different methods for a speech phoneme /m/ taken from a female utterance "him" at SNR = 0 dB.

in terms of the poles of the AR systems in a unified fashion for white noise as well as periodic impulse-train excitations. The ARRC model has been extended for the ARMA system resulting an ARMA ramp cepstrum (ARMARC) model which is also valid for the both type of excitations. The residue-based LS ramp-cepstral fitting scheme that employs the proposed RC model has been presented in order to estimate the AR and MA parameters. The proposed system identification methods can estimate the desired system parameters with sufficient accuracy under noisy environments. The new algorithm using the ARMARC model directly computes both AR and MA parameters of the ARMA systems in accordance with a proper optimization technique even from noise-corrupted observations. On the other hand, the two-stage (TS) algorithm uses a residual signal to estimate the MA parameters with the help

of an IACF and provides comparatively better MA estimates than that obtained by the two-step (OS) method. Even though *a priori* knowledge of the pole strength and location is not necessary for the proposed method to perform, such a knowledge, if available, could easily be incorporated in the proposed scheme to reduce the search range.

In the proposed ramp-cepstrum methods, the system order is assumed to be known. As mentioned in Chapter 2, in the case of real-life data (i.e., the data in the presence of noise) a bad choice of the model order would affect the estimation accuracy of the parameters for all the methods. In the proposed ramp-cepstrum model-fitting based AR estimation algorithm, the poles are determined successively, a simple one or a pair of complex ones at a time. Thus, the accuracy of estimated poles obtained by the proposed method is not at all affected if the chosen AR model order (P') is less than the true order (P). However, if $P' > P$, the accuracy of the first P poles will not be affected. For the ARMA system identification, the AR estimation part works independent of the MA estimation part, and the poles are determined successively, a simple one or a pair of complex ones at a time. Thus, the accuracy of at least those poles which are estimated using the proposed method is not affected if the model order is chosen wrongly as (P', Q') for an ARMA(P, Q) model for the case when $Q' < P' < P$.

As described in Chapter 2, similar to the case of ARMAC mode-fitting based method, the computational complexity of the proposed ramp-cepstrum model-fitting based methods depends on the number of search points used in the RBLS optimization process. As mentioned earlier, in the proposed methods, a significant reduction in the computational complexity is obtained based on the neighborhood search of the initial frequency estimates and a two-stage coarse and fine search scheme for the

magnitude estimation. The computational time required by the proposed methods is also found quite reasonable for practical applications where the objective is to achieve an accurate estimation of the system parameters at very low levels of SNR.

From an extensive simulation on different synthetic systems, it has been shown that the proposed method is able to estimate the system parameters with sufficient accuracy and consistency for signals, at very low levels of SNR, in the presence of noise. As an application of the proposed method the vocal-tract system identification in the presence of white noise is performed using both AR and ARMA model demonstrating a superior estimation performance.

Some of the distinctive features of the proposed ramp-cepstrum method of system identification reinforced by the experimental results can be summarized as follows.

1. In the proposed identification methods, the advantageous features of the correlation and cepstrum have been utilized to obtain a better estimation accuracy.

2. Because of the noise-compensation from the ACF of the residual signal, a better accuracy in the MA parameter estimation is obtained.

3. In the proposed method, the input excitation power need not be assumed to be known.

4. It is capable of handling the problem of AR or ARMA system identification under severe noisy conditions for both white noise and impulse-train input excitations. This feature makes the method readily applicable to speech signals.

5. The proposed method estimates the system parameters with guaranteed stability.

Chapter 4

Ramp Cosine Cepstrum Model Based System Identification

4.1 Introduction

In this Chapter, new schemes based on the ramp cosine cepstrum (RCC) of the one-sided autocorrelation function (OSACF) are proposed for the parameter estimation of minimum-phase AR and ARMA systems under low levels of SNR. Two ramp cosine cepstrum models valid for both white noise and periodic impulse-train excitations are proposed for the OSACF of AR and ARMA signals [111], [112]. In order to estimate the AR parameters of the AR or ARMA system from the noise-corrupted output observations with a guaranteed system stability, the residue-based least-squares optimization algorithm as described in Chapter 2 is employed in conjunction with the RCC model. Since the proposed RCC model-fitting approach makes use of the attractive features of both the correlation- and cepstral-domain representations of the signal, a more accurate parameter estimation is obtained in the presence of heavy noise. In order to estimate the MA parameters of the ARMA system, two different approaches are developed in a manner similar to that proposed in Chapter 3. In the first approach, a one-step (OS) or direct scheme is proposed to estimate the MA

parameters using the RCC model together with the RBLS algorithm. In the second approach, a two-stage (TS) scheme is proposed where a residual signal is first obtained by filtering the observed data via the estimated AR parameters. Then, the MA parameters are estimated from the residual signal by using the noise-compensated scheme proposed in Chapter 2. For the purpose of implementation, the discrete cosine transform (DCT), which can efficiently handle the phase unwrapping problem and offer computational advantages over the discrete Fourier transform, is employed. Extensive simulations are carried out on synthetic AR and ARMA systems of different orders in the presence of noise. Simulation results demonstrate quite a satisfactory identification performance even for an SNR of -5 dB. To illustrate the suitability of the proposed technique in practical applications, the human vocal-tract system identification is also carried out using natural speech signals.

The rest of the chapter is organized as follows. In Section 4.2, the AR system identification method is described. In this section, first, an RCC model for the OSACF 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. An AR parameter estimation scheme under noisy conditions is then described. Simulation results on different synthetic AR systems and natural speech signal are presented in Section 4.3. In Section 4.4, the proposed ARMA system identification methods is described. Here, the RCC model of the OSACF of ARMA signal is first developed and then two different approaches for the estimation of the system parameters are introduced. The estimation performance of the proposed ARMA system identification method is demonstrated in Section 4.5 through simulations for both synthetic and natural speech signals. Finally, in Section 4.6, some key features of the proposed methods are summarized with concluding remarks.

4.2 AR System Identification

4.2.1 Problem Statement

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

$$x(n) = -\sum_{k=1}^{P} a_k x(n-k) + u(n) \tag{4.1}$$

where $u(n)$ and $x(n)$ are, respectively, the excitation and the response of the AR system, $\{a_k\}$ the AR parameters to be estimated, and P the system order assumed to be known. Note that when the system order is unknown, different standard techniques, available in the literature [33], [98], can be employed to estimate the order. The system output in (4.1) can be considered as a convolution of the input $u(n)$ and the impulse-response $h(n)$ of the system, represented as

$$x(n) = h(n) * u(n) \tag{4.2}$$

The transfer function of the AR(P) system described by (4.1) can be written as

$$H(z) = \frac{1}{A(z)} = \frac{1}{\prod_{k=1}^{P}(1 - p_k z^{-1})} \tag{4.3}$$

where $A(z) = \sum_{k=1}^{P} a_k z^{-k}$ 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. It is assumed that the AR process is wide sense stationary. In most of the system identification problems, $u(n)$ is modeled to be a stationary zero-mean white Gaussian noise with an unknown variance σ_u^2. For some practical applications, such as speech signal processing, seismology, and communication, however, the excitation may have other forms [7], [8], [100]. For example, in speech signal processing, a periodic impulse-train is often used as an

excitation of the vocal-tract system [7], [8]. As such, 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 [7], [8]. For an N-point real sequence $\{s(n)\}_{n=0}^{N-1}$, in general, the cepstrum of $s(n)$ can be defined as [113]

$$\gamma_s(n) = \mathcal{T}^{-1}[\ln[\mathcal{T}[s(n)]]] \tag{4.4}$$

where $\mathcal{T}[\cdot]$ and $\mathcal{T}^{-1}[\cdot]$, respectively, represent a transform and its inverse operator. As usual, the natural logarithm is used. 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

$$\ln[S(z)] = \ln[|S(z)|] + j\angle S(z) \tag{4.5}$$

Note that, the sequence $s(n)$ has a real, stable, and uniquely defined cepstrum if $\ln[S(z)]$ has a convergence power series representation. This implies that $\ln[S(z)]$ should be an analytical function within a region of convergence including the unit circle. Hence, both $\ln[|S(z)|]$ and $\arg[S(z)]$ must be a continuous function of ω. When $S(z)$ does not have zeros on the unit circle, the continuity of $\ln[|S(z)|]$ is guaranteed. However, since, a numerical computation of (4.5) provides only the principal or wrapped phase, a phase unwrapping algorithm is necessary to restore the phase continuity [8], [114]–[116].

In the current system identification problem, the system response $x(n)$, as described in (4.2), is a convolution of the input and the impulse-response of the system, and its cepstrum, which can be expressed as a sum of the two corresponding cep-

strums, is given by

$$\gamma_x(n) = \gamma_h(n) + \gamma_u(n) \tag{4.6}$$

where $\gamma_h(n)$ and $\gamma_u(n)$ are the cepstrum of the impulse response $h(n)$ and the input signal $u(n)$, respectively. Utilizing such an advantage of homomorphic deconvolution, cepstrum domain methods have been proposed for system identification in [73], [86], [117], [118]. For example, in [73], in order to estimate the AR parameters, a mean-squared error minimization involving (4.6) is used by employing the Cholesky decomposition. However, as mentioned in [73], 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

$$y(n) = x(n) + v(n) \tag{4.7}$$

where $v(n)$ is assumed to be a zero mean stationary process and is independent of $u(n)$. In [119], the behavior of the cepstral coefficients affected by 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 consider a more common and critical situation where only noisy observations are available. Using the definition given in (4.4), the complex cepstrum of $y(n)$ can be expressed as

$$\begin{aligned}\gamma_y(n) &= \mathcal{T}^{-1}\{\ln[\mathcal{T}[x(n)]]\} + \mathcal{T}^{-1}\left\{\ln\left[1 + \frac{\mathcal{T}[v(n)]}{\mathcal{T}[x(n)]}\right]\right\} \\ &= \gamma_x(n) + \gamma_w(n)\end{aligned} \tag{4.8}$$

where the term $\gamma_w(n)$ arises from the presence of $v(n)$, and vanishes in its absence. It can be observed from (4.8) that in the cepstral domain the effect of noise is additive. In order to estimate the AR system parameters from $\gamma_x(n)$, the effect of $\gamma_w(n)$ 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 [71], [119]. 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 and carrying out a model-fitting based least-squares optimization. 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. Another advantage of the DCT over the DFT is the computational efficiency in dealing with real signals.

4.2.2 Proposed Ramp Cosine Cepstrum (RCC) Model of One-Sided ACF of AR signal

In the cepstral analysis, cepstral coefficients are, generally, computed from an observed signal or from an estimate of its non-parametric power spectral density (PSD) [11], [33]. 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

$$\psi_x(\tau) = \begin{cases} \phi_x(\tau), & \tau > 0 \\ 0.5\phi_x(\tau), & \tau = 0 \\ 0, & \tau < 0 \end{cases} \tag{4.9}$$

where $\phi_x(\tau)$ is the conventional two-sided ACF of $x(n)$ which, in general, is estimated as [33]

$$\phi_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} x(n)x(n+|\tau|), \; 0 \leq |\tau| < N \tag{4.10}$$

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. Since, $\phi_x(\tau)$ is a symmetric two-sided sequence, the OSACF $\psi_x(\tau)$ is related to $\phi_x(\tau)$ as

$$\phi_x(\tau) = \psi_x(\tau) + \psi_x(-\tau). \tag{4.11}$$

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 [108].

Taking the z-transform of the both sides of (4.11) results in

$$\Phi_x(z) = \Psi_x(z) + \Psi_x(1/z) \tag{4.12}$$

The Fourier domain representation of (4.11) is given by

$$\mathcal{F}[\phi_x(\tau)] = 2\Re e[\mathcal{F}[\psi_x(\tau)]] \tag{4.13}$$

where $\mathcal{F}[\cdot]$ represents the Fourier transform and the operator $\Re e[\cdot]$ gives the real part of a complex number. As we are interested to perform cepstrum domain computation with $\psi_x(\tau)$, the relation in (4.13) favors the use of the cosine transform, which is the real part of the Fourier transform. As discussed in Section 4.2.1, the introduction of cosine transform not only provides advantage for its realization, it also helps in overcoming the phase unwrapping problem. The Fourier cosine transform is the real

part of the full complex Fourier transform and is denoted as $\mathcal{F}_c[\cdot]$. Thus, the Fourier cosine transform of a real signal $\{\psi_x(\tau)\}_{\tau=0}^{N-1}$ can be written as

$$\Psi_x^c(\omega) = \mathcal{F}_c[\psi_x(\tau)] = \Re e[\mathcal{F}[\psi_x(\tau)]] = \sum_{\tau=0}^{N-1} \psi_x(\tau) cos\omega\tau \tag{4.14}$$

From (4.4) and (4.14), one can define the cosine cepstrum of a real signal $\{\psi_x(m)\}_{m=0}^{N-1}$ as

$$c_{\psi_x}(m) = \mathcal{F}_c^{-1}[\ln[\mathcal{F}_c[\psi_x(m)]]] \tag{4.15}$$

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

$$\mathcal{F}_c^{-1}[\Psi_x^c(\omega)] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Psi_x^c(\omega) cos\omega m \; d\omega \tag{4.16}$$

In the following, we will develop a ramp cosine cepstrum model for the estimation of the AR parameters under the white Gaussian noise and a periodic impulse-train excitations. To this end, we first show that the cosine cepstrum $c_{\psi_x}(m)$ can be expressed in terms of the system poles. Using (4.13) and (4.14), $c_{\psi_x}(m)$ in (4.15) can be expressed as

$$c_{\psi_x}(m) = \mathcal{F}_c^{-1}[\ln[\mathcal{F}[\phi_x(m)]]] + \mathcal{F}_c^{-1}\left[\ln\left[\frac{1}{2}\right]\right] \tag{4.17}$$

Here $\mathcal{F}[\phi_x(m)] = \Phi_x(\omega)$ is by definition the PSD of the real signal $x(n)$, and it can be shown that $\Phi_x(\omega)$ is real, even, and non-negative. From (4.2), PSD of the output $x(n)$ for a linear time-invariant system with a transfer function $H(z)$ can be expressed as

$$\Phi_x(z) = H(z)H(z^{-1})\Phi_u(z) = |H(z)|^2 \Phi_u(z) \tag{4.18}$$

where $\Phi_u(z)$ is the PSD of the input signal. Using (4.18), $c_{\psi_x}(m)$ in (4.17) can be written as

$$\begin{aligned} c_{\psi_x}(m) &= \mathcal{F}_c^{-1}[\ln[H(\omega)]] + \mathcal{F}_c^{-1}[\ln[H(-\omega)]] \\ &+\mathcal{F}_c^{-1}[\ln[\Phi_w(\omega)]] + \mathcal{F}_c^{-1}\left[\ln\left[\frac{1}{2}\right]\right] \end{aligned} \tag{4.19}$$

It is observed from (4.19) 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 (4.19) individually. From (4.3), $\ln[H(z)]$ can be expanded as

$$\ln[H(z)] = -\sum_{i=1}^{P} \ln\left(1 - p_i z^{-1}\right) = \sum_{i=1}^{P}\sum_{n=1}^{\infty} \frac{p_i^n}{n} z^{-n} \tag{4.20}$$

Using (4.16), the inverse cosine transform of $\ln[\mathcal{F}[h(n)]]$, with $h(n)$ being real and minimum phase, can be obtained as

$$\begin{aligned} \mathcal{F}_c^{-1}[ln[H(\omega)]] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{i=1}^{P}\sum_{m=1}^{\infty} \frac{p_i^m}{m} e^{-j\omega m} cos\omega m \; d\omega \\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi} \sum_{i=1}^{P}\sum_{m=1}^{\infty} \frac{p_i^m}{m}[1 + \cos 2\omega m] \; d\omega \\ &\quad -j\frac{1}{4\pi}\int_{-\pi}^{\pi} \sum_{i=1}^{P}\sum_{m=1}^{\infty} \frac{p_i^m}{m}[\sin 2\omega m] \; d\omega \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{m=1}^{\infty} \frac{p_i^m}{m}, \; m > 0 \end{aligned} \tag{4.21}$$

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

$$\begin{aligned} \mathcal{F}_c^{-1}[ln[H(-\omega)]] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{i=1}^{P}\sum_{m=1}^{\infty} \frac{p_i^m}{m} e^{j\omega m} cos\omega m \; d\omega \\ &= \frac{1}{2}\sum_{i=1}^{P}\sum_{m=1}^{\infty} \frac{p_i^m}{m}, \; m > 0 \end{aligned} \tag{4.22}$$

It is observed from (4.16) that for a constant value of $\Psi_x^c(\omega)$, $\mathcal{F}_c^{-1}[\Psi_x^c(\omega)] = 0$ for all $m > 0$. Thus for $m > 0$, the last term on the right side of (4.19) vanishes. Let us

now consider the remaining third term of (4.19) that depends on the characteristics of the input excitation $w(n)$. In the following we consider separately (a) the white Gaussian noise and (b) a periodic impulse-train as an input excitation.

(a) White Noise Excitation

For a zero mean white Gaussian noise with a variance σ_w^2, $\Phi_w(z) = \sigma_w^2$. Thus the third term on the right side of (4.19) reduces to

$$\mathcal{F}_c^{-1}[\ln[\Phi_w(\omega)]] = \mathcal{F}_c^{-1}[ln[\sigma_w^2]] = 0, \quad m > 0 \tag{4.23}$$

Hence, for the white noise excitation, the cosine cepstrum $c_{\psi_x}(m)$ in (4.19) can finally be expressed as

$$c_{\psi_x}(m) = \sum_{i=1}^{M} \frac{p_i^m}{m}, \quad m > 0 \tag{4.24}$$

It can be observed from this equation that $c_{\psi_x}(m)$ decays rapidly with increasing m, thus making it difficult to use $c_{\psi_x}(m)$ 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

$$\chi_x(m) = mc_{\psi_x}(m) = \sum_{i=1}^{P} p_i^m, \quad m > 0 \tag{4.25}$$

Since, the poles in a system could appear as real or as complex conjugate pair, (4.25) can be rewritten as

$$\chi_x(m) \quad = \quad \sum_{i=1}^{\kappa} \alpha(\omega_i) r_i^m \cos(\omega_i m), m > 0 \tag{4.26}$$

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, and $\alpha(\omega_i)$ is given by

$$\alpha(\omega_i) \quad = \quad \begin{cases} 1, & \omega_i = 0 \text{ or } \omega_i = \pi \\ 2, & 0 < \omega_i < \pi \end{cases} \tag{4.27}$$

is introduced to distinguish real and complex poles. The model given by (4.26) 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.

(b) 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 $m > 0$, 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 (4.19), that is now denoted as $\hat{c}_{\phi_{wi}}(m)$, on $c_{\phi_x}(m)$.

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

$$u_i(n) = \sum_{k=0}^{\mu-1} \delta(n - kT), \tag{4.28}$$

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 (4.10), an estimate of the ACF of $u_i(n)$ is obtained as

$$\phi_{u_i}(\tau) = \frac{1}{N} \sum_{l=0}^{\mu-1} (\mu - l)\delta(|\tau| - lT), \ 0 \leq |\tau| < N \tag{4.29}$$

It is observed from (4.29) that $\phi_{u_i}(\tau)$ 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 $0 \leq |\tau| < N$. Thus, $\phi_{u_i}(\tau)$ can be expressed alternately as

$$\phi_{u_i}(\tau) = \begin{cases} f\left(\frac{\tau}{T}\right), & |\tau| = 0, T, 2T, \ldots, (\mu - 1)T \\ 0, & \text{otherwise} \end{cases} \tag{4.30}$$

where

$$f(\tau) = \begin{cases} \dfrac{\mu - |\tau|}{N}, & |\tau| \leq \mu - 1 \\ 0, & \text{otherwise} \end{cases} \tag{4.31}$$

Note that $f(\tau)$ is an even symmetric triangular sequence and from (4.30) and (4.31), it is evident that $f(\tau)$ can be obtained by down-sampling $\phi_{u_i}(\tau)$ with a factor T. Thus the z transform of $\phi_{u_i}(\tau)$ can be expressed as

$$\Phi_{u_i}(z) = F(z^T) \tag{4.32}$$

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

$$F(z) = \frac{1}{N} \frac{(z^\mu - 1)^2}{z^{\mu-1}(z-1)^2}, \quad z \neq 0, 1 \tag{4.33}$$

Based on the relation between $\phi_{u_i}(\tau)$ and $f(\tau)$, as described in (4.30), (4.31) and (4.32), it can be shown that

$$\hat{c}_{\phi_{ui}}(m) = \begin{cases} \hat{c}_f\left(\dfrac{m}{T}\right), & m = 0, T, 2T, \ldots, (\mu - 1)T \\ 0, & \text{otherwise} \end{cases} \tag{4.34}$$

where

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

It is evident from (4.34) that $\hat{c}_{\phi_{ui}}(m)$ assumes non-zero values at $m = 0$ and at integral multiples of T for $m \geq 0$. Thus, the third term on the right side of (4.19) reduces to

$$\mathcal{F}_c^{-1}[\ln[\Phi_{u_i}(\omega)]] = 0, \quad 0 < m < T \tag{4.35}$$

Note that the RCC given by (4.25) for the white noise excitation can be modified for the impulse-train excitation as

$$\chi_x(m) = mc_{\psi_x}(m) = \sum_{i=1}^{P} p_i^m, \quad 0 < m < T \tag{4.36}$$

From (4.25) and (4.36), 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 $0 < m < T$.

4.2.3 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 reduces 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 (4.15) can be implemented as follows

$$\begin{aligned} \chi_x(m) &= mc_{\psi_x}(m),\ m > 0 \\ c_{\psi_x}(m) &= IDCT[\ln(DCT[\psi_x(m)])],\ m > 0 \end{aligned} \tag{4.37}$$

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

$$DCT[\psi_x(n)] = \Psi_x(k) = \varsigma(k) \sum_{n=0}^{N-1} \psi_x(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right)$$
$$k = 0, 1, \cdots, N-1 \tag{4.38}$$

$$IDCT[\Psi_x(k)] = \psi_x(n) = \sum_{k=0}^{N-1} \varsigma(k)\Psi_x(k) \cos\left(\frac{(2n+1)k\pi}{2N}\right)$$
$$n = 0, 1, \cdots, N-1 \tag{4.39}$$

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

$$\varsigma(k) = \begin{cases} \sqrt{1/N}, & \text{for } k = 0 \\ \sqrt{2/N}, & \text{for } k = 1, 2, \ldots, N-1 \end{cases} \tag{4.40}$$

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 (4.37) can be easily carried out on and (4.37) can be expressed as

$$\varsigma_x(n) = \Re e\left[IDCT[\ln|\Psi_x(k)| + j\pi\xi]\right], n > 0 \tag{4.41}$$

where

$$\xi = \begin{cases} 0, & \text{if } \Psi_x(k) \geq 0 \\ -1, & \text{if } \Psi_x(k) < 0 \end{cases} \tag{4.42}$$

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 [8], [114]–[116] need to be used, since the phase in this case has no longer binary values.

4.2.4 Ramp Cosine Cepstral Fitting: Residue Based Least-Squares Minimization

In the presence of noise, the observed signal gets heavily corrupted especially when the signal-to-noise ratio (SNR) is very low. In Section 4.2.1, 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 [108]. 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

$$\phi_y(\tau) = \phi_x(\tau) + \phi_n(\tau) \tag{4.43}$$

where

$$\phi_n(\tau) = \phi_v(\tau) + \phi_{xv}(\tau) + \phi_{vx}(\tau) \tag{4.44}$$

Here $\phi_v(\tau)$ is the ACF of noise $v(n)$, and $\phi_{xv}(\tau)$ and $\phi_{vx}(\tau)$ are crosscorrelation terms. 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 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

$$\psi_y(\tau) = \begin{cases} \phi_y(\tau), & \tau > 0 \\ 0.5\phi_y(\tau), & \tau = 0 \\ 0, & \tau < 0 \end{cases} \tag{4.45}$$

From (4.43) and (4.45), the OSACF of $y(n)$ can be written as

$$\psi_y(\tau) = \psi_x(\tau) + \psi_n(\tau) \tag{4.46}$$

where $\psi_n(\tau)$ indicates an effect of noise on $\psi_y(\tau)$ and it can be expressed in a form similar to that of $\psi_x(\tau)$ given by (4.9). Thus, in the presence of noise, the cosine cepstrum of $\psi_y(\tau)$ can be expressed as

$$c_{\psi_y}(m) = \mathcal{F}_c^{-1}[\ln[\mathcal{F}_c[\psi_y(m)]]] = c_{\psi_x}(m) + c_{\psi_n}(m), \quad n > 0 \tag{4.47}$$

where

$$c_{\psi_n}(m) = \mathcal{F}_c^{-1}\left\{\ln\left[1 + \frac{\mathcal{F}_c[\psi_n(m)]}{\mathcal{F}_c[\psi_x(m)]}\right]\right\} \tag{4.48}$$

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

$$\chi_y(n) = \chi_x(n) + \chi_\epsilon(n), \quad n > 0 \tag{4.49}$$

Here, $\chi_\epsilon(n)$ is the error introduced due to the noise. It is to be noted that the effect of noise, which is additive in the observed noisy signal given by (4.7) or its ACF given by (4.43), can also be treated as additive in the proposed ramp cosine cepstrum domain. Now, the RCC model derived in Section 4.2.2 can be used in (4.49) 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, one way of reducing

the effect of noise is to compensate the amount of noise from $\phi_y(0)$. In this case the noise variance must be known or estimated. However, in most practical applications, the noise variance is *a priori* not known or very difficult to obtain. One possible way of estimating the noise variance is to utilize the noise-only data, if available. This technique is employed in speech signal processing when the observed noisy data contains noise-only (pause) segments. In the proposed identification scheme, a more general noisy environment is considered where it is assumed that the noise variance is unknown and noise-only data is not available. 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, by excluding the zero lag one may reduce the effect of noise. However, in this case, the average power of the observed data $y(n)$ will be removed. 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, instead of discarding the zero lag altogether. The process can efficiently suppress the level of $c_{\psi_n}(m)$ while leaving the shape of $c_{\psi_y}(m)$ similar to that of $c_{\psi_x}(m)$.

As discussed in the previous sub-section following (4.49) that a ramp cosine cepstral fitting approach can be employed to determine the RCC model parameters from the RCC of the OSACF of noisy observations. Then, the AR parameters can be obtained from the RCC model parameters $\{r_i\}$ and $\{\omega_i\}$. Each of the κ component terms in (4.26) 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 $\Re_{l-1}(n)$ and the lth

component of the RCC model, that is

$$J_l = \sum_{n=1}^{N_c} |\Re_{l-1}(n) - \alpha(\omega_l) r_l^n \cos(\omega_l n)|^2 \,,\; l = 1, 2, \ldots, \kappa \tag{4.50}$$

where the residual function is updated as follows

$$\begin{aligned} \Re_0(n) &= \chi_y(n) \\ \Re_l(n) &= \Re_{l-1}(n) - \alpha(\omega_l) r_l^n \cos(\omega_l n),\; l = 1, \ldots, \kappa - 1 \end{aligned} \tag{4.51}$$

Note that $\{r_l\}$ and $\{\omega_l\}$ are independent variables and α depends on $\{\omega_l\}$ as given by (4.27) and it is not an independent variable. We would like to find the optimal solution for $\{r_l\}$ and $\{\omega_l\}$ by a search algorithm based on the computation of (4.50) and (4.51) in which different sets of values for $\{r_l\}$ and $\{\omega_l\}$ in a bounded region are tested for a possible solution. In order to estimate each of the κ such pairs, J_l given by (4.50) is computed for different trial values of $\{r_l\}$ and $\{\omega_l\}$. The values of one pair (r_l, ω_l) corresponding to the global minimum of J_l are selected as the estimate of the desired poles. It can be observed from (4.51) that, in order to determine the l-th residual function $\Re_l(n)$, already computed values of $\{r_l\}$ and $\{\omega_l\}$ are utilized. Proceeding in this manner, the AR parameters can be determined using (4.3) once all the P 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. Note that instead of the entire RCC model with all κ constituent terms, each such term is estimated sequentially in (4.50). This is done with a view to convert a multi-dimensional optimization problem into a set of two-dimensional optimization scheme which makes the problem much simpler. In order to reduce the computational burden, a two-step search algorithm can be used. In

the first step, only 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 $\{r_l\}$, and in the second step, a fine-search is carried out around the initial estimate with a higher resolution to obtain a more accurate estimate. Once the magnitudes and angles of the desired poles are obtained, the AR parameters can be computed using the relation given by (4.3).

4.3 Simulation Results on AR System Identification

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 improved least-squares algorithm with a faster convergence (ILSF) or ILSD method [62], signal/sub-space Yule-Walker (SSYW) method [58], and modified least-squares YW (MLSYW) [33] method.

4.3.1 Results on Synthetic AR Systems

(a) White Noise Excitation: A noisy signal is generated according to (4.1) and (4.7) with $N = 4,000$ and $\sigma_u^2 = 1$, where the variance of the white Gaussian noise σ_v^2

is appropriately set based on a specified level of SNR defined as

$$\text{SNR} = 10\log_{10}\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} v(n)^2} \quad \text{dB} \tag{4.52}$$

From the noisy observations, first, the OSACF $\psi_y(\tau)$ is computed using (4.45) and (4.10). As given in (4.37) 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 described in Section 4.2.4. 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 even for systems with a very fast decaying autocorrelations. The initial estimates of ω_l is 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 P unknown AR coefficients, the number of RCC samples to be considered in the model-fitting operation should be higher than P. In our experiment, the number of RCC samples is taken as $N_c = 10P$.

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 $|\phi_y(1)|/\phi_y(0)$. An experiment consists of $N_T = 100$ independent trials to find the means and variances of the estimated AR parameters. The experiments are conducted for 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, i.e., the true value (SDT), and (4) the average sum-squared error (ASSE) given by

$$\text{ASSE} = \frac{1}{N_T P} \sum_{m=1}^{N_T} \sum_{k=1}^{P} [\hat{a}_k(m) - a_k]^2 \tag{4.53}$$

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

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 4.1 and 4.2 show the estimation results for the AR(3) and AR(4) systems at an SNR level of -5 dB, respectively. The AR(3) system contains a real pole and a pair of complex conjugate pole, and the AR(4) system contains two real poles and a pair of complex conjugate pole. 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 4.1 shows that at SNR $= -5$ dB, when the other methods fail to identify the system, the proposed method successfully estimates the parameters with quite accurately. It is seen from Table 4.2, although some of the other methods provide an acceptable performances, the estimation accuracy achieved by the proposed method is much higher. It is seen from these tables that the proposed method exhibits superior estimation performance with respect to all the four performance indices at such a

Table 4.1: 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 4.2: 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

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.

Fig. 4.1 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 4.1. It is observed from Fig. 4.1 that the ILSF and the SSYW methods give estimation accuracy comparable to that provided by the proposed method for SNR levels above

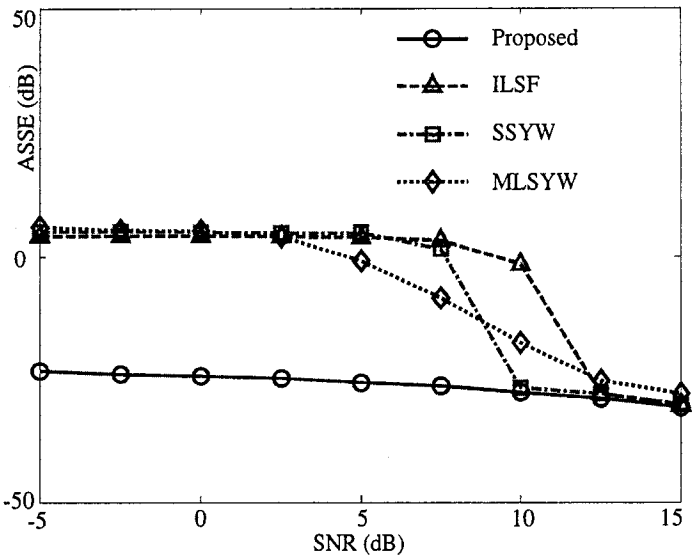


Figure 4.1: Effect of noise level on the ASSE for a white noise excited system.

10 dB. However, the proposed method performs significantly better for levels of SNR as low as -5 dB.

Fig. 4.2 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 given by

$$a_k = \{1, -3.2229, 5.2862, -5.0095, 2.7875, -0.7362\}.$$

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.

(b) 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 (4.28) with a known value of T. We

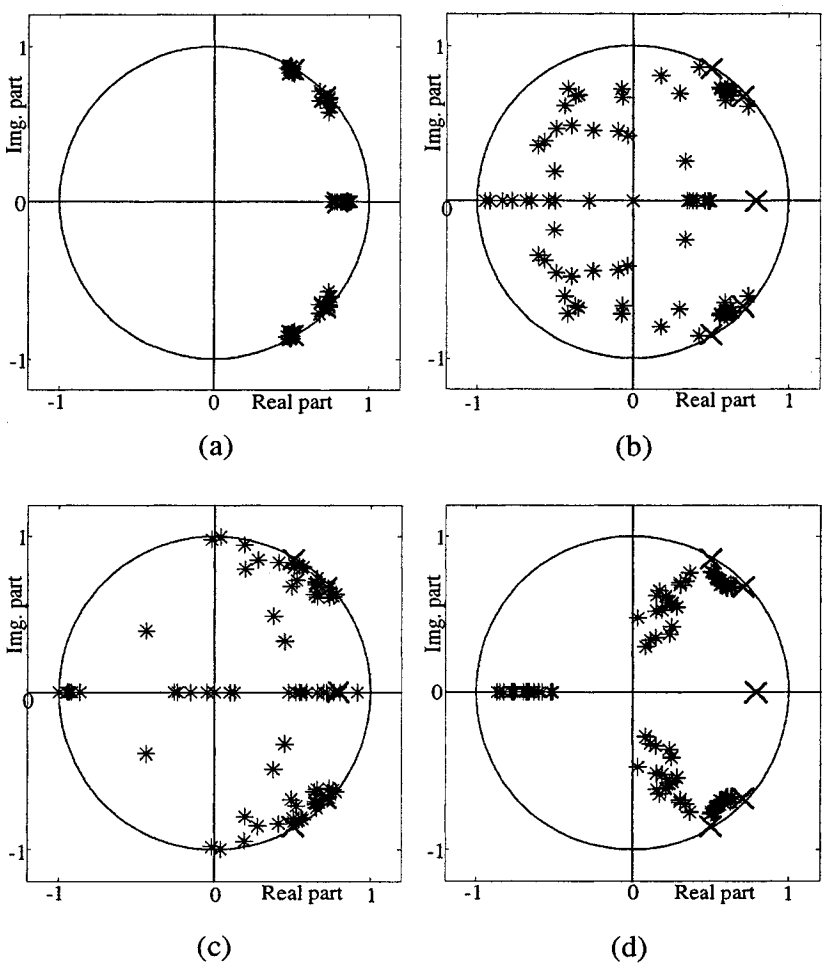


Figure 4.2: Superimposed pole plot of AR(5) system at SNR = -5 dB. $\times$: true poles and $*$: estimated poles. (a) Proposed, (b) ILSF, (c) SSYW, and (d) MLSYW method.

Table 4.3: 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
a_1	−2.6770	−2.6816	−0.9615	−1.0776	−1.0588
		(±0.0311)	(±1.0657)	(±0.6890)	(±0.0962)
		(±0.0290)	(±2.0196)	(±1.7414)	(±1.6211)
a_2	2.5894	2.5644	−0.2284	0.1767	−0.1307
		(±0.0702)	(±1.7218)	(±0.7101)	(±0.1728)
		(±0.0671)	(±3.3022)	(±2.5150)	(±2.7256)
a_3	−0.8970	−0.8732	0.4773	0.2489	0.5269
		(±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

choose the number of RCC samples less than T; thus, $N_c = min(T-1, 10P)$. A noisy AR signal is generated according to (4.1) and (4.7) with $N = 4,000$. The simulations are carried out for $N_T = 100$ independent trials and the results averaged.

Tables 4.3 and 4.4 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 an 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 Fig. 4.3. 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.

Table 4.4: 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	0.3845	0.3719	0.1483
		(±0.0432)	(±0.2824)	(±0.3122)	(±0.4145)
		(±0.0456)	(±0.2914)	(±0.3134)	(±0.4225)
a_2	−0.0100	−0.0591	0.0151	0.0247	0.0608
		(±0.0501)	(±0.0705)	(±0.0607)	(±0.0615)
		(±0.0540)	(±0.0743)	(±0.0699)	(±0.0938)
a_3	−0.7853	−0.7483	−0.8134	−0.8428	−0.7973
		(±0.0651)	(±0.0602)	(±0.0402)	(±0.0387)
		(±0.0730)	(±0.0664)	(±0.0701)	(±0.0407)
a_4	−0.5999	−0.5568	−0.4196	−0.4663	−0.2965
		(±0.0658)	(±0.2953)	(±0.2885)	(±0.1842)
		(±0.0660)	(±0.3107)	(±0.2992)	(±0.3549)
ASSE (dB)		−22.84	−11.35	−10.71	−8.27

4.3.2 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, non-parametric 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 [120], [121] 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

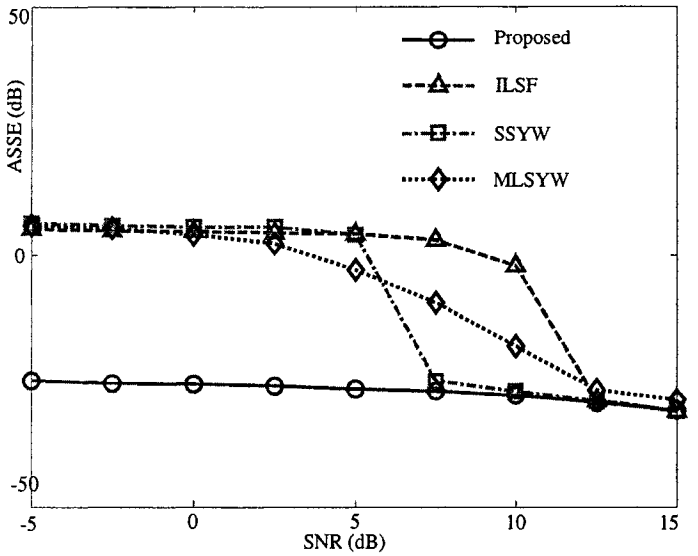
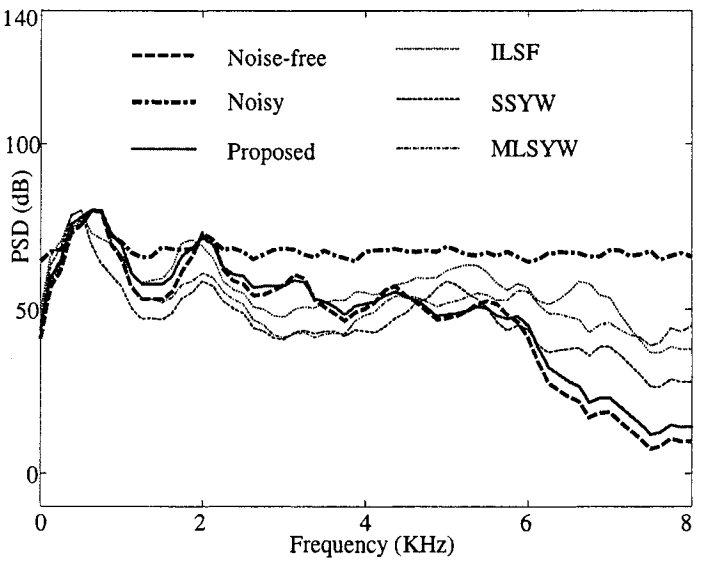


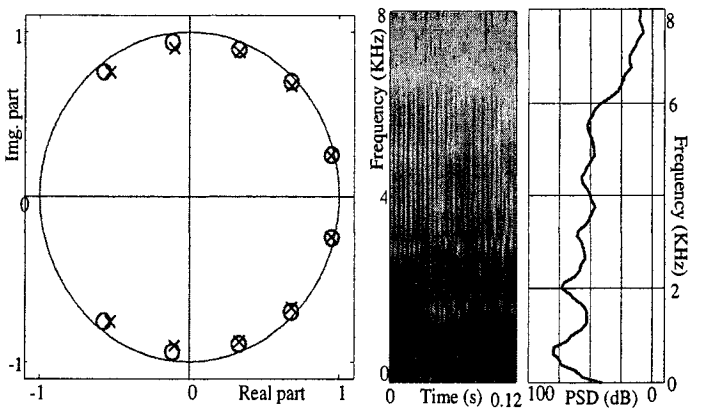
Figure 4.3: Effect of noise level on the ASSE for an impulse-train excited system.

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 [7]. For computing synthesized speech signals by different methods, same excitation signal is used for a particular phoneme. It is to be mentioned that the synthesized sounds obtained by different methods were also played back in order to test the subjective quality. 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 is compared with that obtained in a noise-free condition by using the MLSYW method. Quality of the synthesized sounds obtained by the proposed method under such a noisy condition were found far superior than that obtained by the other methods. Fig. 4.4(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 /*a*/ of the word '*Rob*' uttered by a male speaker. The estimated pole locations of the vocal-tract system are averaged over 20 independent noisy realizations and used to obtain the synthesized speech. We choose the number of RCC samples N_c less than the pitch period T; thus, $N_c = min(T-1, 10P)$. According to the general behavior of the vocal tract parameter, r_l is searched in the range $[0.8, 0.99]$ [93]. The search range for ω_l can be narrowed down based on the knowledge of the pole locations of a particular phoneme [7], [93]. 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 Fig 4.4(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 Fig. 4.4(b) along with the noise-free estimates obtained by the MLSYW method. In Fig. 4.4(b), the noise-free wide-band spectrogram and the noise-free non-parametric 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 multi-talker babble noise (multiple background competing speakers) taken from the Noisex92 database [122]. In Fig. 4.5(a), the results obtained at an SNR of -5 dB for a naturally spoken sound /ε/ of the word '*head*' uttered by a female speaker are presented. The multiplicity of speakers produces a flatter short-term

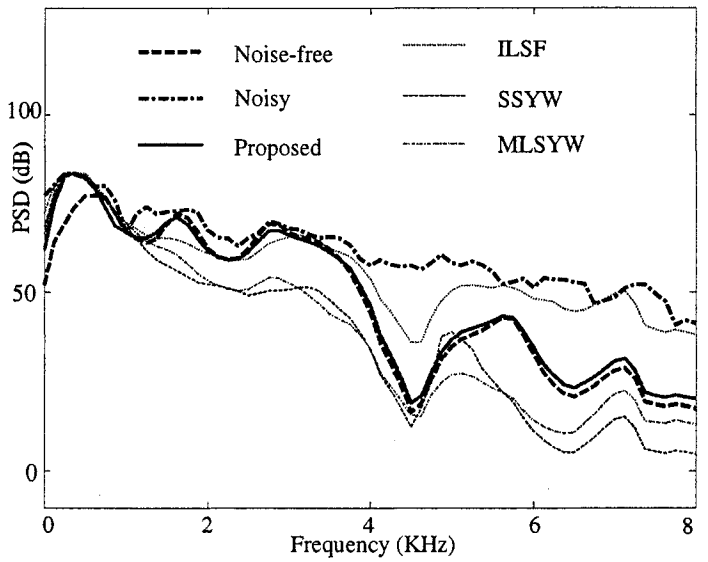


(a)

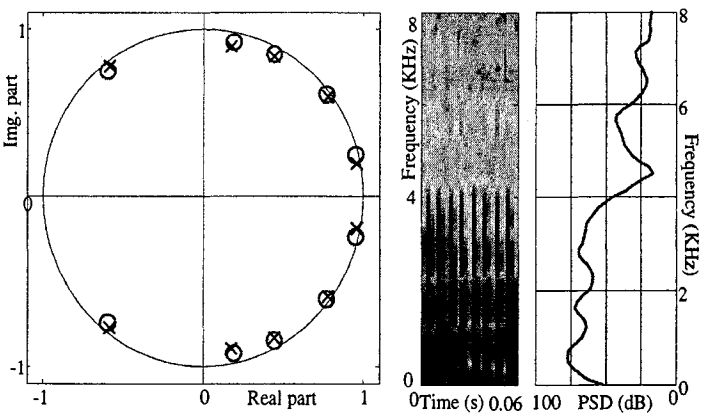


(b)

Figure 4.4: Estimation results for a natural speech phoneme /ε/ 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.



(a)



(b)

Figure 4.5: Estimation results for a natural speech phoneme /a/ in the presence of a multi-talker babble noise at SNR = -5 dB. (a) PSD obtained by using different methods, (b) Average estimated poles ($\times$) obtained from noise-corrupted speech by using the proposed method along with the noise-free estimates ($\circ$) obtained by the MLSYW method, wide-band spectrogram of the noise-free speech, and noise-free PSD.

spectrum which has greater spectral and temporal modulation than a white Gaussian noise. It is observed from Fig. 4.5(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 Fig. 4.5 (b). In this figure the estimated average poles along with the noise-free pole estimates, the wide-band spectrogram, and the non-parametric PSD are shown. Fig. 4.5 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.

4.4 ARMA System Identification

In this section, a ramp cosine cepstrum model corresponding to the ARMA system is developed. It is to be mentioned that in Chapter 3, based on the principle of derivation of the ARRC model, an ARMARC model is derived and implemented for the ARMA system identification. It can be realized from the detail derivation of the ARRCC model as described in the previous section that, an RCC model corresponding to the ARMA system needs to be derived based on a similar procedure by taking into account some necessary changes due to the presence of zeros apart from the poles. An ARMA system identification scheme is then developed based on the model-fitting approach.

4.4.1 Problem Formulation

According to the system description given in Chapter 2, the input $u(n)$ and the output $x(n)$ of a causal stable and LTI ARMA (P, Q) system are characterized by

$$\sum_{i=0}^{P} a_i x(n-i) = \sum_{j=0}^{Q} b_j u(n-j) \tag{4.54}$$

where a_i and b_j are the AR and MA parameters with $a_0 = 1$ and $b_0 = 1$, and P and Q are the orders of the ARMA model, which are assumed to be knownThe corresponding system transfer function is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{\prod_{j=1}^{Q}(1 - z_j z^{-1})}{\prod_{k=1}^{P}(1 - p_k z^{-1})} = \sum_{k=1}^{P} \frac{\eta_k}{1 - p_k z^{-1}} \tag{4.55}$$

where $A(z) = \sum_{k=1}^{P} a_k z^{-k}$ and $B(z) = 1 + \sum_{j=1}^{Q} b_j z^{-j}$ are, respectively, the AR and MA polynomial, p_k's and z_j's denote, respectively, the poles and the zeros of the ARMA system. It is assumed that all poles and zeros are of the first-order and the the ARMA(P, Q) process is minimum phase and stationary.

From an OSACF $\psi_x(m)$ of the ARMA signal $x(n)$, according to the definition given in (4.15), the cosine cepstrum of $\psi_x(m)$ can be expressed as

$$\begin{aligned} c_{\psi_x}(m) &= \mathcal{F}_c^{-1}[\ln[H(\omega)]] + \mathcal{F}_c^{-1}[\ln[H(-\omega)]] \\ &+ \mathcal{F}_c^{-1}[\ln[\Phi_w(\omega)]] + \mathcal{F}_c^{-1}\left[\ln\left[\frac{1}{2}\right]\right] \end{aligned} \tag{4.56}$$

For the case of an ARMA system described by (4.55), $\ln[H(z)]$ can be expanded as

$$\begin{aligned} \ln[H(z)] &= -\sum_{i=1}^{P} \ln\left(1 - p_i z^{-1}\right) + \sum_{j=1}^{Q} \ln\left(1 - z_j z^{-1}\right) \\ &= \sum_{i=1}^{P}\sum_{n=1}^{\infty} \frac{p_i^n}{n} z^{-n} - \sum_{j=1}^{Q}\sum_{n=1}^{\infty} \frac{z_j^n}{n} z^{-n} \end{aligned} \tag{4.57}$$

It can easily be observed from (4.57) that, in comparison to (4.20) for the case of ARMA system additional terms appear due to the presence of system zeros. In order to develop an identification algorithm for the ARMA system, our objective is now to determine all four parts of (4.56) and then develop a ramp-cosine cepstrum model.

4.4.2 Proposed RCC Model of OSACF of ARMA Signal and Parameter Estimation

Based on the definition given in (4.16), with the impulse response of the ARMA system $h(n)$ being real and minimum phase, $\mathcal{F}_c^{-1}[ln[H(\omega)]]$ in (4.56) can be obtained as

$$\begin{aligned}
\mathcal{F}_c^{-1}[ln[H(\omega)]] &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\sum_{i=1}^{P}\sum_{m=1}^{\infty}\frac{p_i^m}{m}e^{-j\omega m} - \sum_{j=1}^{Q}\sum_{m=1}^{\infty}\frac{z_j^m}{m}e^{-j\omega m}\right] cos\omega m\ d\omega \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\sum_{i=1}^{P}\sum_{m=1}^{\infty}\frac{p_i^m}{m} - \sum_{j=1}^{Q}\sum_{m=1}^{\infty}\frac{z_j^m}{m}\right](1+\cos 2\omega m)\ d\omega \\
&\quad -j\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\sum_{i=1}^{P}\sum_{m=1}^{\infty}\frac{p_i^m}{m} - \sum_{j=1}^{Q}\sum_{m=1}^{\infty}\frac{z_j^m}{m}\right](\sin 2\omega m)\ d\omega \\
&= \frac{1}{2}\sum_{i=1}^{P}\sum_{m=1}^{\infty}\frac{p_i^m}{m} - \frac{1}{2}\sum_{j=1}^{Q}\sum_{m=1}^{\infty}\frac{p_j^m}{m},\ m>0 \qquad (4.58)
\end{aligned}$$

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

$$\begin{aligned}
\mathcal{F}_c^{-1}[ln[H(-\omega)]] &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\sum_{i=1}^{P}\sum_{m=1}^{\infty}\frac{p_i^m}{m}e^{j\omega m} - \sum_{j=1}^{Q}\sum_{m=1}^{\infty}\frac{z_j^m}{m}e^{j\omega m}\right] cos\omega m\ d\omega \\
&= \frac{1}{2}\sum_{i=1}^{M}\sum_{m=1}^{\infty}\frac{p_i^m}{m} - \frac{1}{2}\sum_{j=1}^{Q}\sum_{m=1}^{\infty}\frac{p_j^m}{m},\ m>0 \qquad (4.59)
\end{aligned}$$

It can be shown in a similar way as stated above for the AR system that, for the white noise excitation, the cosine cepstrum $c_{\psi_x}(m)$ can be expressed as

$$c_{\psi_x}(m) = \sum_{i=1}^{P}\frac{p_i^m}{m} - \sum_{j=1}^{Q}\frac{z_j^m}{m},\quad m>0 \qquad (4.60)$$

A ramp cosine cepstrum (RCC) for the OSACF of $x(n)$ is thus obtained as

$$\chi_x(m) = mc_{\psi_x}(m) = \sum_{i=1}^{P}p_i^m - \sum_{j=1}^{Q}z_j^m,\quad m>0 \qquad (4.61)$$

For a periodic impulse-train excitation $\{u_i(n)\}_{n=0}^{N-1}$ with period T, as derived above for the AR system, it can be shown that $\hat{c}_{\phi_{ui}}(m)$ assumes non-zero values at $m = 0$

and at integral multiples of T for $m \geq 0$. Thus, $c_{\psi_x}(m)$ can be written as

$$\chi_x(m) = mc_{\psi_x}(m) = \sum_{i=1}^{P} p_i^m - \sum_{j=1}^{Q} p_j^m, \quad 0 < m < T \tag{4.62}$$

which is same as (4.61) except the range is now $0 < m < T$ instead of $m > 0$. For a real-valued $x(n)$, complex poles (zeros) will always appear as conjugate pairs. In (4.61), the complex pole (zero) pairs and real poles (zeros) will each contribute one decaying exponential, which can be written as a decaying cosine function, yielding

$$\chi_x(m) = \sum_{i=1}^{\kappa_P} \alpha(\omega_{p_i}) r_{p_i}^m \cos(\omega_{p_i} m) - \sum_{j=1}^{\kappa_Q} \alpha(\omega_{z_j}) r_{z_j}^m \cos(\omega_{z_j} m), m > 0 \tag{4.63}$$

where $\kappa_P(\kappa_Q)$ represents the number of real poles (zeros) plus the number of complex conjugate pole (zero) pairs, r_{p_i} and ω_{p_i} are, respectively, the magnitude and the argument of the ith pole p_i, and r_{z_j} and r_{z_j} and ω_{z_j} are, respectively, the magnitude and the argument of the jth zero z_j. In (4.63), $\alpha(\omega)$ is introduced to distinguish the real and complex zeros and poles, namely, $\alpha(\omega) = 1$ if $\omega = 0$ or π, otherwise $\alpha(\omega) = 2$. As a result, (4.63) can be expressed as

$$\chi_x(m) = \sum_{k=1}^{\kappa' = \kappa_P + \kappa_Q} \alpha(\omega_k) r_k^m \cos(\omega_k m), \ m > 0 \tag{4.64}$$

where $\alpha(\omega_k)$ can be written as

$$\alpha(\omega_k) = \begin{cases} (-1)^\eta, & \omega_k = 0 \text{ or } \omega_k = \pi \\ (-1)^\eta 2, & 0 < \omega_k < \pi \end{cases} \tag{4.65}$$

with η given by

$$\eta = \begin{cases} 0, & k \leq \kappa_P \\ 1, & k > \kappa_P \end{cases} \tag{4.66}$$

The model given by (4.64) is termed as the ARMA ramp cosine cepstrum (ARMARCC) model for the OSACF of $x(n)$ which will be used to form an objective

function for the LS fitting. Note that in case of impulse-train excitation as we mentioned in (4.64), $0 < m < T$ has to be considered instead of $m > 0$. For the purpose of implementation, like the case of AR system identification, DCT-IDCT can be employed.

A ramp cosine cepstral fitting approach can be employed to determine the RCC model parameters $\{r_k\}$ and $\{\omega_k\}$ from the RCC of the OSACF of noisy observations. Each of the κ' component functions in (4.64) is estimated sequentially from N_c nonzero instances of $\chi_y(m)$, where $N_c < T$ for the impulse-train excitation. The objective function can be formulated in a similar fashion as it is done for the AR system identification. Therefore, the total squared error between the $(l-1)$th residual function $\Re_{l-1}(n)$ and the lth component of the model is given by

$$J_l = \sum_{m=1}^{N_c} \left| \Re_{l-1}(m) - \alpha(\omega_l) r_l^m \cos(\omega_l m) \right|^2, l = 1, \ldots, \kappa' \tag{4.67}$$

where the residual function is updated as follows

$$\begin{aligned} \Re_0(m) &= \psi_y(m) \\ \Re_l(m) &= \Re_{l-1}(m) - \alpha(\omega_l) r_l^n \cos(\omega_l m),\ l = 1, \ldots, \kappa' - 1 \end{aligned} \tag{4.68}$$

The optimal solution for $\{r_l\}$ and $\{\omega_l\}$ are then obtained through the search algorithm described in Chapter 3. The two-step search algorithm described before can be also employed to reduce the computational burden. In the first step, only 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 $\{r_l\}$, and in the second step, a fine-search is carried out around the initial estimate with a higher resolution to obtain a more accurate estimate. Once the magnitudes and angles of the desired poles are obtained, the AR parameters can be computed using the relation given by (4.55). In the proposed

search scheme, restricting the search range of r_l within the stable region inherently guarantees the stability of the system.

As an alternate of the above one-step (OS) method, a two-step (TS) MA parameter estimation algorithm proposed in Chapter 2 can easily be employed once the AR parameters are obtained using the ARMA ramp-cepstral least-square minimization. In this case, first, a residual signal is obtained by filtering the noisy observed signal via the estimated AR parameters. Then the zeros are computed from the IACF corresponding to the noise-compensated ACF of the residual signal using the RBLS algorithm.

The residue-based LS ramp-cepstral fitting scheme that employs the proposed RC model has been presented in order to estimate the AR and MA parameters. The proposed system identification methods can estimate the desired system parameters with sufficient accuracy under noisy environments. Using the ARMARC model, the proposed OS algorithm directly computes both AR and MA parameters of the ARMA systems in accordance with a proper optimization technique even from noise-corrupted observations. On the other hand, the TS algorithm uses a residual signal to estimate the MA parameters with the help of an IACF and provides comparatively better MA estimates than that obtained by the OS method. Even though *a priori* knowledge of the pole strength and location is not necessary for the proposed method to perform, such a knowledge, if available, could easily be incorporated in the proposed scheme to reduce the search range. From an extensive simulation on different synthetic systems, it has been shown that the proposed method is able to estimate the system parameters with sufficient accuracy and consistency for signals, at very low levels of SNR, in the presence of noise. As an application of the proposed method the vocal-tract system identification in the presence of white noise is performed using both AR and ARMA

model demonstrating a superior estimation performance.

4.5 Simulation Results on ARMA System Identification

In this section, the estimation performance of the proposed OS and TS methods under noisy conditions is evaluated and compared with that of the ARMA cepstrum recursion (ACR) method [45], and the order-selective Durbin's (OSD) method [33] considered for the purpose of comparison in Chapter 3. An ARMA signal is generated according to (4.7) and (4.54) with a data length of $N = 4,000$ and excitation variance $\sigma_u^2 = 1$, where the noise variance σ_v^2 is appropriately determined according to a specified level of SNR defined in (4.52). The RCC model parameters are determined using the RBLS optimization algorithm where the search ranges and resolutions for ω_l and r_l and initial estimates of ω_l are kept same as that used for the RCC model based AR system identification. Each experiment contains $N_T = 100$ independent trials and we compute the estimation mean, SDM, SDT, and ASSE as defined before.

In Table 4.5, estimation results for an ARMA$(4, 3)$ system is presented at SNR -5 dB. In this table, the last row gives the true parameter values. The estimated values of the corresponding parameters obtained by different methods including the proposed OS and TS methods are given in the preceding four rows. The corresponding values for the SDM and SDT are shown within the parentheses below the each estimated parameter value. It is found that both TS and OS methods provide quite an accurate parameter estimation performance in the presence of noise in comparison to the other methods. The small values of standard deviations SDM and SDT obtained by the proposed techniques indicate respectively a high estimation consistency and accuracy. It is observed from our experimentation that the performance of the TS method is

Table 4.5: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited ARMA(4, 3) system at SNR = −5 dB

Methods	Estimated parameters							ASSE
	a_1	a_2	a_3	a_4	b_1	b_2	b_3	(dB)
Proposed	−2.5771	3.33559	−2.2429	0.7645	−2.0272	1.8558	−0.6569	−17.41
(TS)	(±0.1064)	(±0.1974)	(±0.1928)	(±0.0949)	(±0.0863)	(±0.1036)	(±0.0545)	
	(±0.1076)	(±0.1935)	(±0.1970)	(±0.1027)	(±0.1080)	(±0.1043)	(±0.0552)	
Proposed	−2.5771	3.3559	−2.2429	0.7645	−1.9875	1.8019	−0.6145	−14.48
(OS)	(±0.1064)	(±0.1974)	(±0.1928)	(±0.0949)	(±0.2122)	(±0.2072)	(±0.0902)	
	(±0.1076)	(±0.1935)	(±0.1970)	(±0.1027)	(±0.2159)	(±0.2083)	(±0.0928)	
OSD	−0.6300	0.0914	0.1891	−0.1116	−0.4909	0.1258	0.0701	5.88
	(±0.1185)	(±0.1580)	(±0.1384)	(±0.0922)	(±0.5152)	(±0.4191)	(±0.2625)	
	(±1.9685)	(±3.2514)	(±2.3931)	(±0.6263)	(±1.6821)	(±1.7683)	(±0.7646)	
ACR	−0.6300	0.0914	0.1891	−0.1116	−0.5234	0.0068	0.1076	5.87
	(±0.1185)	(±0.1580)	(±0.1384)	(±0.0922)	(±0.1188)	(±0.1462)	(±0.1247)	
	(±1.9685)	(±3.2514)	(±2.3931)	(±0.6263)	(±1.5732)	(±1.8428)	(±0.7659)	
True	−2.5950	3.3390	−2.2000	0.7310	−2.0922	1.8438	−0.6480	

slightly better than that of the OS method which supports the explanation given before.

Fig. 4.6 presents the ASSE values as a function of the level of SNR for all four methods for an ARMA(3, 2) system with true parameters given below

$$a_i = \{1, -2.5712, 2.5218, -0.9460\}$$

$$b_j = \{1, -1.6909, 0.81\}.$$

From this figure it can be observed that the proposed TS and OS methods exhibit consistently better estimation performance over the range of SNR considered in comparison to the other two methods. The estimation error obtained by the OSD and ACR methods are very high at low levels of SNR. The proposed TS and OS methods exhibit excellent performance even at a low level of SNR.

Figs. 4.7 and 4.8 present the average estimated poles and zeros of an ARMA(6, 4) system obtained by the proposed TS and OS methods at SNR = −5 dB. The

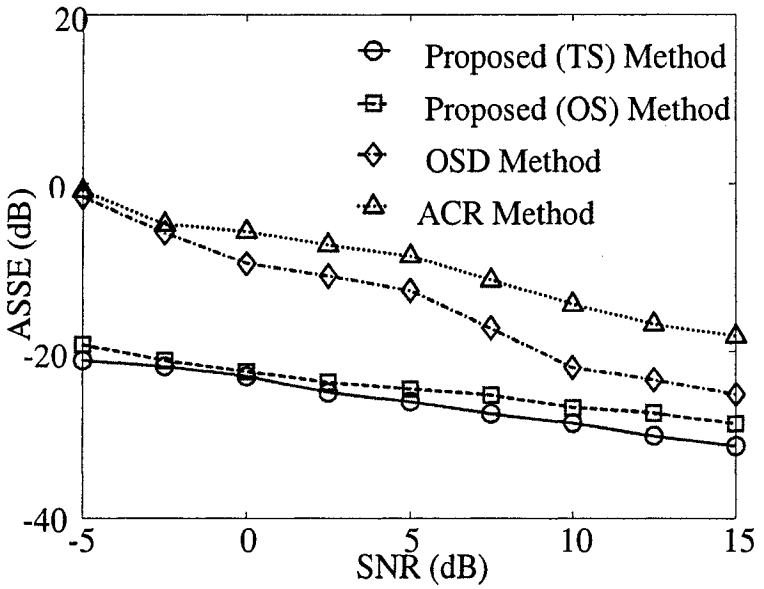


Figure 4.6: Effect of SNR on the estimation accuracy for the white noise excited ARMA(3, 2) system.

ARMA(6, 4) system considered in this figure has following parameters

$$a_i = \{1, -1.2174, 1.2225, -1.0537, 1.1167, -1.0253, 0.85\}$$

$$b_j = \{1, -1.1354, 0.9399, -0.4974, 0.2198\}.$$

In this figure the true poles and zeros are also included for the purpose of comparison. From this figure it can be observed that the proposed methods are able estimate the poles and zeros quite accurately at a very low level of SNR.

Identification performance of the proposed methods are also tested for the impulse-train excited systems in the presence of noise. A periodic impulse train with a known value of T is generated in a same manner as done for the RCC based AR system identification. Fig. 4.9 shows, the ASSE values are plotted as a function of levels of SNR obtained by different methods for the same ARMA(3, 2) system as considered

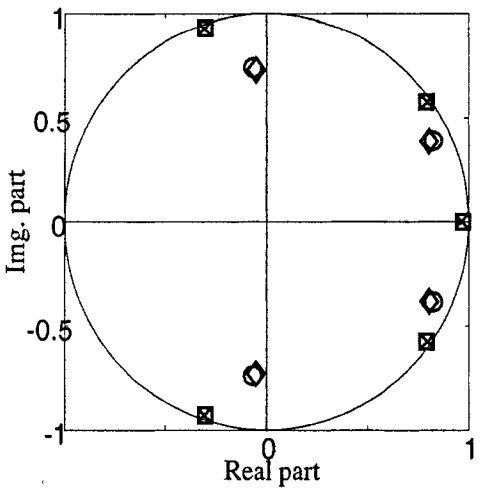


Figure 4.7: Superimposed pole-zero plot of ARMA(5, 4) system obtained by the TS method at SNR = -5 dB; (Poles: $\times$: true, $\square$: proposed; Zeros: O: true, $\diamondsuit$:proposed).

in Fig. 4.6 under an impulse-train excitation. Here the period T of the impulse train is chosen as 68 and number of ramp-cepstral instances $M_c = min(T/2, 10P)$. It is clearly observed that the proposed method provides a significant performance at a very low SNR while the performance of the other two methods deteriorates at low levels of SNR.

As an application of the proposed RCC based ARMA system identification method, the estimation of a vocal tract system parameters is performed using natural speech signal in a similar manner as it was presented in Chapter 2. For the purpose of testing some English nasal sounds (voiced phonemes) from the TIMIT standard database are used. No pre-filtering is performed in order to observe the accuracy of the pole-zero estimation over the entire range of frequency. The PSD of the synthesized speech signal obtained by different methods from the estimated vocal-tract system parameters is compared with that of the noise-free natural speech. The synthesized sounds obtained by different methods were also played back and the quality of the synthe-

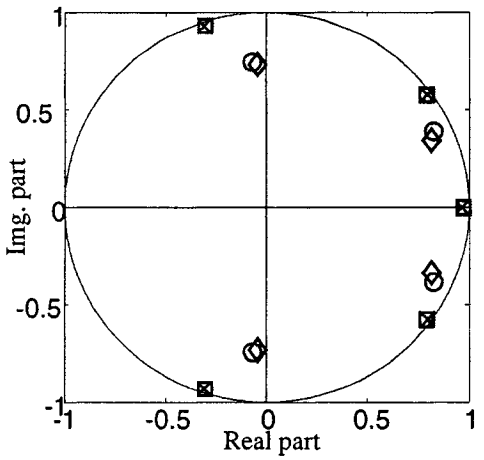


Figure 4.8: Superimposed pole-zero plot of ARMA(5, 4) system obtained by the OS method at SNR = -5 dB; (Poles: $\times$: true, $\square$: proposed; Zeros: O: true, $\diamondsuit$:proposed).

sized sounds obtained by the proposed method under such a noisy condition were found far superior than that obtained by the other methods. Fig. 4.10 shows the PSD of the synthesized speech obtained by using different methods along with that of the noise-free speech, and that obtained from one of the 20 noisy signals under a white Gaussian noise of SNR = -5 dB. Here we consider an ARMA$(12, 6)$ model, for a naturally spoken female nasal sound $/m/$ of the word "him". It is seen from Fig. 4.10 even at a very low level of SNR, the proposed method exhibits a superior identification performance in comparison to other methods.

4.6 Conclusion

In this chapter, new techniques for the parameter estimation of AR and ARMA systems from noise-corrupted output observations, have been proposed. Comprehensive and accurate ramp cosine cepstrum (RCC) models for the one-sided ACF of an AR and ARMA signals, valid for both white noise and periodic impulse-train excitations,

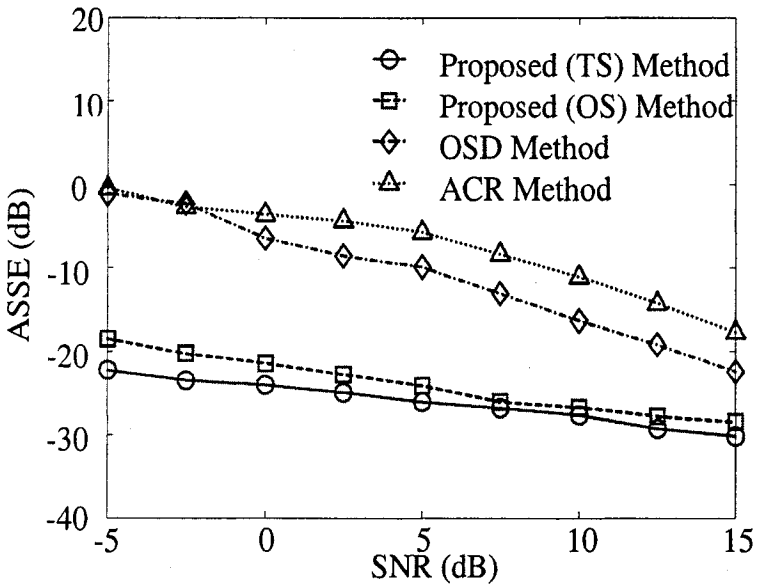


Figure 4.9: Effect of SNR on the estimation accuracy for the impulse-train excited ARMA(3, 2) system.

have been developed in a unified fashion for the identification of the AR or ARMA systems. A residue-based least-squares ramp cosine cepstral fitting scheme employing the RCC model has been presented for a more accurate estimation of the AR parameters of the AR or ARMA system. 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 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.

In the proposed ramp cosine cepstrum methods, the system order is assumed to be known. As mentioned in Chapters 2 and 3, in the case of real-life data (i.e., the data in the presence of noise) a bad choice of the model order would affect the estimation accuracy of the parameters for all the methods. Like the ramp-cepstrum

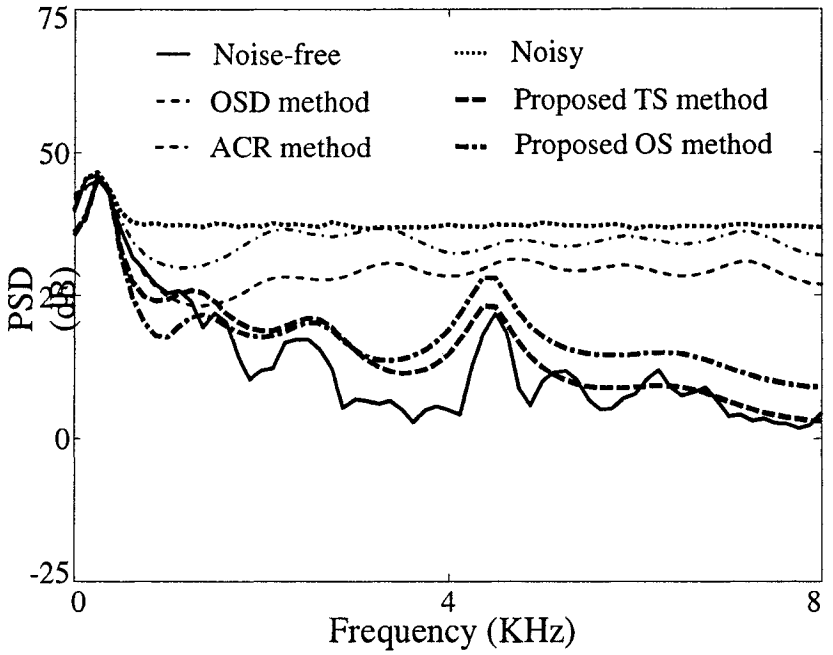


Figure 4.10: PSD obtained by using different methods for a speech phoneme /m/ taken from a female utterance "him" at SNR = −5 dB.

method, in the proposed ramp cosine cepstrum model-fitting based AR estimation algorithm, the poles are determined successively, a simple one or a pair of complex ones at a time. Thus, the accuracy of the estimated poles obtained by the proposed method is not at all affected if the chosen AR model order (P') is less than the true order (P). However, if $P' > P$, the accuracy of the first P poles will not be affected. For the ARMA system identification, the AR estimation part works independent of the MA estimation part, and the poles are determined successively, a simple one or a pair of complex ones at a time. Thus, the accuracy of at least those poles which are estimated using the proposed method is not affected if the model order is chosen wrongly as (P', Q') for an ARMA(P, Q) model for the case when $Q' < P' < P$.

As described in Chapter 3, similar to the case of ramp-cepstrum mode-fitting based

method, the computational complexity of the proposed ramp cosine cepstrum model-fitting based methods depends on the number of search points used in the RBLS optimization process. As mentioned earlier, in the proposed methods, a significant reduction in the computational complexity is obtained based on the neighborhood search of the initial frequency estimates and a two-stage coarse and fine search scheme for the magnitude estimation. The computational time required by the proposed methods is also found quite reasonable for practical applications where the objective is to achieve an accurate estimation of the system parameters at very low levels of SNR.

A comprehensive simulation study performed on different AR and ARMA systems has demonstrated that the proposed method is sufficiently accurate and consistent in estimating the system parameters 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 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 corrupted by white or real-life babble noise.

Some of the distinctive features of the proposed ramp cosine cepstrum method of system identification reinforced by the experimental results can be summarized as follows.

1. The proposed method combines the attractive features of the correlation- and cepstral-domain system identifications in order to obtain a better estimation accuracy.

2. For the purpose of implementation, unlike the ramp-cepstrum method, in the

proposed ramp cosine cepstrum method, the DCT is employed which is capable of handling the phase unwrapping problem and offers computational advantages over the DFT.

3. Because of the noise-compensation from the ACF of the residual signal, a better accuracy in the MA parameter estimation is obtained.

4. In the proposed method, the input excitation power need not be assumed to be known.

5. It is capable of handling the problem of AR or ARMA system identification under severe noisy conditions for both white noise and impulse-train input excitations. This feature makes the method readily applicable to speech signals.

6. The proposed method estimates the system parameters with guaranteed stability.

Chapter 5

ARMA System Identification Based on Noise-Compensation in the Correlation Domain

5.1 Introduction

For the estimation of the AR parameters of the ARMA system, correlation-domain Yule-Walker (YW) methods are most widely employed. In a noisy environment, the performance of these methods gets degraded. One way to overcome this problem, is the zero-lag compensation of the ACF in the low-order YW methods,which require a prior estimate of the noise variance that is not available in most of the practical applications. An alternate solution could be zero-lag exclusion, but it would affect the parameter estimation accuracy in the case of finite data length, since the lower lags of the ACF are more reliable in terms of information contents than the higher lags. Our target is to perform the identification task accurately even in noise, without discarding the zero lag and at the same time without requiring a prior knowledge of the noise variance.

In this chapter, a new scheme for the identification of minimum-phase autoregressive moving average (ARMA) systems from noise-corrupted observations is pre-

sented [123]–[125]. From the autocorrelation function (ACF) of the observed data, by exploiting the characteristics of the zero lag, a set of equations containing the lower lags of the ACF is established to form a quadratic eigenvalue problem in order to estimate simultaneously the AR parameters and the observation noise variance [126]. In the proposed identification technique, both the white noise and the periodic impulse-train excitations are considered for the purpose of practical applications. In order to estimate the MA parameters, first, a residual signal is obtained by filtering the noisy observations via the estimated AR parameters. A noise-subtraction algorithm is proposed utilizing the estimated noise variance together with the AR parameters to reduce the effect of noise from the ACF of the residual signal. The MA parameters are then estimated by using the spectral factorization corresponding to the noise-compensated ACF of the residual signal [127], [128]. In order to demonstrate the effectiveness of the proposed method, extensive simulations are performed by considering synthetic ARMA systems of different orders in the presence of additive white noise and the results are compared with those from some of the existing methods. Computer simulations demonstrating superior identification results in terms of estimation accuracy and consistency even under a heavy noisy condition are conducted. Simulation results for the identification of a human vocal-tract system using natural speech signals are also provided showing a superior estimation performance as per system pole location.

The rest of the chapter is organized as follows. In Section 5.2, we demonstrate the problem of AR and MA parameter estimation in the presence of noise. Section 5.3 presents the proposed methodology to estimate the AR parameters of the ARMA system first for the case of white noise excitation and then for the periodic impulse train excitation. In Section 5.4, we describe a technique to estimate the MA parame-

ters of the ARMA system for white noise and the periodic impulse train excitations. The performance of the proposed method is demonstrated in Section 5.5 through extensive computer simulations for both synthetic and natural speech signals. Finally, in Section 5.6, the salient features of this investigation are summarized.

5.2 Problem Statement

5.2.1 Estimation of AR Parameters

A causal, stable and LTI ARMA(P,Q) system can be characterized by

$$\sum_{i=0}^{P} a_i x(n-i) = \sum_{j=0}^{Q} b_j u(n-j) \tag{5.1}$$

where $u(n)$ and $x(n)$ are, respectively, the excitation and the response of the system, a_i and b_j the corresponding AR and MA parameters which are real with $a_0 = 1$ and $b_0 = 1$, and P and Q $(P > Q)$ are orders of the ARMA model, which are assumed to be known. In the proposed identification scheme, both white noise and periodic impulse-train excitations are considered. The system output in (5.1) can be written as

$$x(n) = h(n) * u(n) = \sum_{k=0}^{\infty} h(k)u(n-k) \tag{5.2}$$

where the impulse-response $h(n)$ is causal and the ARMA system is assumed to be initially relaxed. The system transfer function is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{\prod_{j=1}^{Q}(1 - z_j z^{-1})}{\prod_{k=1}^{P}(1 - p_k z^{-1})} = \sum_{k=1}^{P} \frac{\eta_k}{1 - p_k z^{-1}} \tag{5.3}$$

where p_k and z_j denote, respectively, the poles and the zeros of the ARMA system, η_k the partial fraction coefficient corresponding to the kth pole.Then the corresponding

system impulse response is readily given by

$$h(n) = \sum_{k=1}^{P} \eta_k p_k^n \tag{5.4}$$

As in most of the ARMA system identification methods, it is assumed that all poles and zeros are of the first-order, no further pole-zero cancelation is possible, and the ARMA(P, Q) process is minimum phase. It is assumed that $x(n)$ is a wide sense stationary and second order ergodic process, which means the time average can be used to replace the statistical expectation.

The autocorrelation function (ACF) of $x(n)$ is defined as [129]

$$r_x(\tau) = E[x(n)x(n-\tau)] \tag{5.5}$$

where τ indicates the ACF lag and $r_x(\tau)$ is even-symmetric with respect to zero lag ($\tau = 0$). Given a finite set of observations $\{x(n)\}_{n=0}^{N-1}$, the ACF of $x(n)$ is commonly estimated as [33]

$$r_x(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} x(n)x(n+|\tau|), \ 0 \leq |\tau| < N \tag{5.6}$$

In the correlation-based system identification methods, generally the first few lags (or commonly termed as lower lags) of $r_x(\tau)$ are used in the estimation of the AR parameters. In practical applications, observation data length N is finite and (5.6) offers an efficient way to obtain an accurate estimate of the ACF defined in (5.5). For notational convenience, same symbols are used for the estimated ACF obtained by (5.6) and the ACF defined in (5.5). Multiplying (5.1) by $x(n-\tau)$ and taking expectation, one can obtain

$$r_x(\tau) = -\sum_{i=1}^{P} a_i r_x(\tau - i) + \sum_{j=0}^{Q} b_j E[u(n-j)x(n-\tau)] \tag{5.7}$$

where the last term can be rewritten as

$$\sum_{j=0}^{Q} b_j E[u(n-j)x(n-\tau)] = \sum_{j=0}^{Q} b_j \sum_{k=0}^{\infty} h(k) r_u(k - \overline{j-\tau}) \tag{5.8}$$

In conventional system identification problems, only the white noise input is considered and in this case, the ACF of the input signal $u(n)$ is given by

$$r_u(\tau) = \sigma_u^2 \delta(\tau) \tag{5.9}$$

Thus, the inner sum in the right hand side of (5.8) reduces to $\sigma_u^2 h(\tau - j)$ which forces the last term in (5.7) to vanish for $\tau > Q$ due to the causality of $h(n)$. As a result, (5.7) can be rewritten as

$$r_x(\tau) = \begin{cases} -\sum_{i=1}^{P} a_i r_x(\tau - i) + \sigma_u^2 \sum_{j=0}^{Q} b_j h(j-\tau), 0 \le \tau \le Q \\ -\sum_{i=1}^{P} a_i r_x(\tau - i), \quad \tau > Q \end{cases} \tag{5.10}$$

The above equation clearly indicates that the AR parameters of the ARMA system can be obtained with an estimate of the ACF of observed data for $\tau > Q$ by solving a set of simple linear equations. This is the basic principle of estimating AR parameters of the ARMA system using the correlation based methods, such as, different variants of Yule-Walker methods [33]. Note that in these methods, as described above, only the white noise excitation is considered and the statistical estimator defined in (5.5) is used to obtain the ACF in the basic formulation. However, in practical applications, an estimate of the ACF is required where the conventional correlation estimator given by (5.6) is used most commonly. In view of dealing with the practical applications, we develop an estimation algorithm for the case of white noise as well as periodic impulse-train excitation based on the conventional estimator, instead of the statistical estimator of the ACF of the observed data.

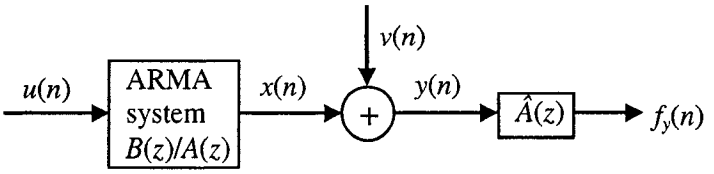


Figure 5.1: Filtering ARMA signal in noise by the estimated AR polynomial.

In the presence of additive white Gaussian noise (AWGN) $v(n)$, the observed signal $y(n)$ can be written as

$$y(n) = x(n) + v(n) \tag{5.11}$$

where $v(n)$ is a stationary process with zero-mean and variance σ_v^2, and is independent of $u(n)$. Estimation of the AR parameters of the ARMA system using Yule-Walker methods results a high variance in the estimation error at a low level of SNR. Especially, the situation is more difficult in the low-order Yule-Walker method where a noise-compensation at the zero lag of the ACF is required, and in conventional methods, it is assumed that the noise-variance is known or *a priori* knowledge about the noise is available. In the high-order Yule-Walker method, in order to avoid the zero lag of the ACF, some equations containing lower lags of the ACF may have to be excluded which eventually affects the estimation accuracy as mentioned earlier that the lower lags of the ACF, in comparison to the higher lags, contain more system information. One of the objectives is to develop an AR parameter estimation algorithm from output observations by utilizing the lower lags of the ACF and a noise-compensation strategy without requiring a prior knowledge of the noise variance.

5.2.2 Estimation of MA Parameters

In order to identify the MA part, one can filter the ARMA signal using the estimated AR polynomial $\hat{A}(z)$ to obtain a residual signal given by

$$f_x(n) = x(n) + \sum_{k=1}^{P} \hat{a}_k x(n-k) \tag{5.12}$$

Durbin's method of MA parameter estimation has been most commonly used in which, first, the intermediate AR system equivalent to the MA residual is identified and then the MA parameters are obtained from these intermediate AR parameters. This method strongly depends on the accuracy of the estimated intermediate AR model parameters. In a noisy environment, identification of the zeros of the ARMA system from output observations is a difficult problem, since the additive noise directly affects the estimation of the system zeros. In this case, the residual signal is obtained by filtering the noisy ARMA signal $y(n)$ with the estimated AR polynomial $\hat{A}(z)$, giving

$$f_y(n) = y(n) + \sum_{k=1}^{P} \hat{a}_k y(n-k) \tag{5.13}$$

Clearly, $f_y(n)$ is also noise-corrupted, making the MA estimation problem difficult. Fig. 5.1 shows the filtering process in the presence of noise $v(n)$. Generally it is assumed that an accurate AR estimate of the ARMA system has already been obtained in the first step, i.e., $\hat{A}(z) \approx A(z)$. Then, (5.13) can be rewritten as

$$f_y(n) = \hat{B}(z)u(n) + \hat{A}(z)v(n) = f_x(n) + f_w(n) \tag{5.14}$$

where $f_x(n)$ is given by (5.12), which is the residual signal with respect to the noise-free ARMA signal $x(n)$, and $f_w(n)$ is the noisy part due to the additive white Gaussian noise $v(n)$. It is evident that $f_w(n)$ is a colored MA(P) sequence, even if the original noise $v(n)$ is white. Thus, $f_y(n)$ can be treated as an MA sequence of order P.

If Durbin's method or any other residual signal based MA estimation methods are employed under noisy condition, the resulting MA estimates will not correspond to $f_x(n)$ due to the presence of $f_w(n)$. The situation becomes worse when the SNR is low. Our target is to propose a new algorithm that first eliminates the noise effect and then determines the MA parameters from the noise-compensated ACF of the residual function.

5.3 Proposed Method of AR Parameter Estimation

5.3.1 White Noise Excitation

It can be observed from (5.10) that the AR parameters of the ARMA system can be obtained with an estimate of the ACF of the observed data for $\tau > Q$ by solving a set of linear equations as given by

$$\begin{bmatrix} r_x(Q) & r_x(Q-1) & \dots & r_x(Q-P+1) \\ r_x(Q+1) & r_x(Q) & \dots & r_x(Q-P+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_x(S-1) & \dots & \dots & r_x(S-P) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} r_x(Q+1) \\ r_x(Q+2) \\ \vdots \\ r_x(S) \end{bmatrix} \tag{5.15}$$

where S governs the number of equations. In general, (5.15) corresponds to an over-determined linear problem, which can be solved by a constrained least-squares algorithm. In the presence of additive white Gaussian noise (AWGN) $v(n)$, the estimated ACF $r_y(\tau)$ of the noisy observation $y(n)$, which is given by (5.11), can be computed using (5.6) as

$$r_y(\tau) = r_x(\tau) + r_v(\tau) + r_\varsigma(\tau) \tag{5.16}$$

where $r_\varsigma(\tau) = r_{xv}(\tau) + r_{vx}(\tau)$ represents the cross-correlation terms. For an AWGN, $r_v(\tau)$ is mostly pronounced at the zero lag and its effect on $r_y(\tau)$ at other lags is negligible. Since $v(n)$ is uncorrelated with $u(n)$ and $x(n)$ is generated from $u(n)$, the effect of the cross-correlation term $r_\varsigma(\tau)$ on $r_y(\tau)$ can be neglected. Thus, one can obtain an estimate of $r_x(\tau)$ from $r_y(\tau)$ using the relation

$$r_x(\tau) = \begin{cases} r_y(\tau) - \lambda, & \tau = 0 \\ r_y(\tau)\ , & \tau \neq 0 \end{cases} \tag{5.17}$$

where $\lambda = r_v(0) = \sigma_v^2$. Clearly, when (5.15) is applied to noisy AR system identification, $r_x(\tau)$ should be replaced by $r_y(\tau)$ except for $\tau = 0$. It is seen from (5.17) that the zero lag of $r_y(\tau)$ could be compensated provided that the noise variance λ can be estimated. In most of the practical applications, however, estimation of the noise variance is a difficult task. Even if it can be estimated, the estimation accuracy could directly affect the solutions of the linear equations (5.15). In the proposed method we are not explicitly estimating the noise variance λ. Instead we will determine the desired AR parameters by formulating a quadratic eigenvalue problem with respect to unknown λ.

Using (5.17), (5.15) can be expressed in the following form

$$\begin{bmatrix} r_y(Q+1) & \cdots & r_y(0)-\lambda & \cdots & r_y(Q+1-P) \\ r_y(Q+2) & \cdots & r_y(1) & \cdots & r_y(Q+2-P) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_y(P) & \cdots & r_y(P-Q+1) & \cdots & r_y(0)-\lambda \\ r_y(P+1) & \cdots & r_y(P-Q+2) & \cdots & r_y(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_y(P+S) & \cdots & r_y(P-Q+S+1) & \cdots & r_y(S) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = 0 \tag{5.18}$$

It can be observed that the first $P - Q$ equations in (5.18) include unknown noise variance λ. Considering $\tau > P$ in (5.18), the zero lag and the unknown noise variance

term can be excluded in the estimation of the AR parameters by deleting the first $P-Q$ equations. However, deletion of these equations would neglect the lower lags of the ACF, which eventually affects the system identification accuracy, since less data remains available to compute the higher lags. We now develop an AR parameter estimation algorithm based on (5.18) without omitting these equations.

By separating the unknown noise variance from the first $P-Q$ equations, (5.18) can be formulated as the following eigenvalue problem,

$$(\mathbf{R}_y - \lambda \mathbf{G})\mathbf{a} = \mathbf{0}_{(P-Q+S)} \tag{5.19}$$

where $\mathbf{0}_{(P-Q+S)}$ is a zero column vector of dimension $(P-Q+S)$, and

$$\mathbf{R}_y = \begin{bmatrix} r_y(Q+1) & \cdots & r_y(0) & \cdots & r_y(Q+1-P) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_y(P) & \cdots & r_y(P-Q+1) & \cdots & r_y(0) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_y(P+S) & \cdots & r_y(P-Q+S+1) & \cdots & r_y(S) \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0} & \mathbf{I}_{(P-Q)} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{a} = [1 \; a_1 \; a_2 \; \ldots \; a_P]^T$$

Note that $\mathbf{R}_y$ and $\mathbf{G}$ are $(P-Q+S) \times (P+1)$ real matrices, and $\mathbf{I}_{P-Q}$ is a $(P-Q) \times (P-Q)$ identity matrix. Considering that there is a total of $p+1$ unknowns $a_i (i = 1, 2, \cdots, P)$ and λ, in addition to the first $(P-Q)$ nonlinear equations in (5.19), the next $S \geq P$ linear equations are chosen. Since $\mathbf{R}_y$ and $\mathbf{G}$ are not square matrices, it is difficult to solve (5.19) by using the standard techniques available for the generalized eigenvalue problem. One way to deal with the rectangular matrix pencil $(\mathbf{R}_y - \lambda \mathbf{G})$ is to convert it into a square one and solve it as a standard generalized eigenvalue problem [130]. However, squaring methods are in general complicated

[130]. As an alternative, pre-multiplying both sides of (5.19) by $(\mathbf{R}_y - \lambda\mathbf{G})^T$, we can convert it into a quadratic eigenvalue problem as given below

$$(\Theta_2\lambda^2 + \Theta_1\lambda + \Theta_0)\mathbf{a} = \mathbf{0}_{(P+1)} \tag{5.20}$$

where

$$\Theta_2 = \mathbf{G}^T\mathbf{G},\ \Theta_1 = -(\mathbf{R}_y^T\mathbf{G} + \mathbf{G}^T\mathbf{R}_y),\ \Theta_0 = \mathbf{R}_y^T\mathbf{R}_y$$

Each of Θ_0, Θ_1 and Θ_2 is real and has a dimension $(P+1)\times(P+1)$. The quadratic eigenvalue problem has been receiving much attention because of its extensive applications in different areas, for example, the dynamic analysis of mechanical systems in acoustics and signal processing in automatic elements of an electric power system [131]. In our quadratic eigenvalue problem (5.20), there are $2(P+1)$ eigenvalues (finite or infinite). However, our goal is to find the eigenvalue giving the noise variance and the corresponding eigenvector which will provide the desired AR parameters. The problem in (5.20) can be solved directly or by employing linearization. Generally speaking, direct methods do not guarantee that the method will converge to the desired eigenvalue. We now solve (5.20) by employing linearization as it transforms (5.20) into an equivalent square linear matrix pencil which can be solved easily by using efficient standard techniques. Substituting $\mathbf{c} = \lambda\mathbf{a}$ into (5.20) yields the following linearized form,

$$\begin{bmatrix} -\Theta_0 & \mathbf{0}_{(P+1)} \\ \mathbf{0}_{(P+1)} & \mathbf{I}_{(P+1)} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \alpha \end{bmatrix} - \lambda \begin{bmatrix} \Theta_1 & \Theta_2 \\ \mathbf{I}_{(P+1)} & \mathbf{0}_{P+1} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \alpha \end{bmatrix} = \mathbf{0}$$

which can be rewritten as

$$(\mathbf{E} - \lambda\mathbf{F})\tilde{\mathbf{a}} = \mathbf{0}_{2(P+1)} \tag{5.21}$$

Note that $\mathbf{E}$ and $\mathbf{F}$ are $2(P+1)\times 2(P+1)$ real square matrices. The problem in (5.21) can be solved by the most widely used QZ algorithm [132]. The QZ method is

numerically stable and can handle the infinite eigenvalue case. Then the eigenvalue λ_{min} having the minimum modulus is chosen as the estimate of the noise variance. The idea of estimating the noise variance by the eigenvalue of minimum magnitude is similar to Pisarenko harmonic retrieval method [33]. In order to obtain estimates of the AR parameters, the eigenvector $\tilde{\mathbf{a}}_{min}$ corresponding to λ_{min} is selected. Note that the first coefficient should be one, which can easily be made by the following normalization

$$\hat{a}_k = \frac{\tilde{a}_{min_k}}{\tilde{a}_{min_1}}, \ k = 2, 3, \ldots, P+1 \tag{5.22}$$

where $\tilde{a}_{min_k}$ is the $k$the component of the chosen eigenvector $\tilde{\mathbf{a}}_{min}$.

As the AR parameters have been estimated together with the estimation of the noise variance in the above process, i.e., the noise compensation has been already implemented implicitly, a better AR estimation accuracy is expected as confirmed in the simulation section.

5.3.2 Impulse-train Excitation

The periodic impulse-train excitation $\{u_i(n)\}_{n=0}^{N-1}$ with a period T can be expressed as

$$u_i(n) = \sum_{m=0}^{\mu-1} \delta(n - mT), \quad \mu = \lceil N/T \rceil \tag{5.23}$$

where $\lceil \zeta \rceil$ represents the smallest integer greater than or equal to ζ, and thus, μ is the total number of impulses in the excitation. Generally, the number of impulses is much smaller than the number of samples within an observation period N, i.e., $\mu \ll N$. The system response $x(n)$ due to a periodic impulse-train excitation is also periodic and can be expressed as

$$x(n) = h(n) * u_i(n) = \sum_{m=0}^{\mu-1} h(n - mT) \tag{5.24}$$

Next, we want to obtain an expression for the ACF $r_x(\tau)$ of $x(n)$. Similar to the white noise excitation, we start with multiplying the characteristic equation (5.1) by $x(n-\tau)$ and taking a time average, yielding

$$\frac{1}{N}\sum_{n=\tau}^{N-1} x(n)x(n-\tau) = -\sum_{i=1}^{P} a_i \left[\frac{1}{N}\sum_{n=\tau}^{N-1} x(n-i)x(n-\tau)\right] + \sum_{j=1}^{Q} b_j \left[\frac{1}{N}\sum_{n=\tau}^{N-1} u_i(n-j)x(n-\tau)\right] \tag{5.25}$$

Also, the ACF defined in (5.6) can be equivalently computed as

$$r_x(\tau) = \frac{1}{N}\sum_{n=\leq|\tau|}^{N-1} x(n)x(n-|\tau|), 0 \leq |\tau| < N \tag{5.26}$$

Using (5.25) and (5.26), one can obtain

$$r_x(\tau) = -\sum_{i=1}^{P} a_i \tilde{r}_x(\tau - i) + r_c(\tau) \tag{5.27}$$

where

$$\sum_{i=1}^{P} a_i \tilde{r}_x(\tau - i) = \sum_{i=1}^{P} a_i r_x(\tau - i) - \sum_{i=1}^{P} a_i \left[\frac{1}{N}\sum_{k=N-i}^{N-1} x(k)x(k-\tau+i)\right] \tag{5.28}$$

and

$$r_c(\tau) = \sum_{j=1}^{Q} b_j \left[\frac{1}{N}\sum_{n=0}^{N-1} u(n-j)x(n-\tau)\right] \tag{5.29}$$

Note that there are only i product terms in the inner sum of the last term of (5.28) as opposed to N terms in $r_x(\tau - i)$, and the maximum value of i is P. Recalling that $P \ll T < N$, $\tilde{r}_x(\tau - i)$ can be approximated with $r_x(\tau - i)$. Thus, (5.27) can be rewritten as

$$r_x(\tau) = -\sum_{i=1}^{P} a_i r_x(\tau - i) + r_c(\tau) \tag{5.30}$$

For a finite length of observation, this is valid for both the white noise and the periodic impulse-train excitations.

In the following, we analyze the effect of $r_c(\tau)$ on $r_x(\tau)$ and show that only a certain number of lags of $r_c(\tau)$ over its period have a significant impact on the computation of $r_x(\tau)$. Clearly, $r_c(\tau)$ is generated through a cross-correlation between some delayed versions of $x(n)$ and $u(n)$. Substituting (5.2) into (5.29) gives

$$r_c(\tau) = \sum_{j=1}^{Q} b_j \left[\frac{1}{N} \sum_{n=0}^{N-1} u(n-j) \left\{ \sum_{k=0}^{\infty} h(k) u(n-\tau-k) \right\} \right] \tag{5.31}$$

Replacing $(\tau + k)$ with l and changing the relevant limit for the sum, we obtain

$$r_c(\tau) = \sum_{j=1}^{Q} b_j \sum_{l=\tau}^{\infty} h(l-\tau) \left[\frac{1}{N} \sum_{n=0}^{N-1} u(n-j) u(n-l) \right] \tag{5.32}$$

Noting that $r_u(\tau) = (1/N) \sum_{n=\tau}^{N-1} u(n)u(n-\tau)$, by replacing further $(n-j)$ with k in the last sum of (5.32), one has

$$r_c(\tau) = \sum_{j=1}^{Q} b_j \sum_{l=\tau}^{\infty} h(l-\tau) \left[r_u(l-j) - r_{ue} \right] \tag{5.33}$$

where the error term r_{ue} can be expressed as

$$r_{ue} = \frac{1}{N} \sum_{k=N-j}^{N-1} u(k) u(k-l+j) \tag{5.34}$$

Since $j \leq Q \ll T$, there are a maximum of Q samples in $u(k)$ which are used in (5.34). Obviously, there can only be at most one impulse in the sum of (5.34). As such, r_{ue} can easily be neglected compared with $r_u(l-j)$. Therefore, $r_c(\tau)$ can be rewritten as

$$r_c(\tau) = \sum_{j=1}^{Q} b_j \sum_{l=\tau}^{\infty} h(l-\tau) r_u(l-j) \tag{5.35}$$

Given the finite duration impulse-train excitation $\{u_i(n)\}_{n=0}^{N-1}$ with period T, its ACF can be calculated as

$$r_{u_i}(\tau) = \frac{1}{N} \sum_{k=0}^{\mu-1} (\mu - k) \delta(\tau - kT), \ 0 \leq \tau < N \tag{5.36}$$

Obviously, $r_u(\tau)$ is an even symmetric decaying function of τ with respect to $\tau = 0$ and has non-zero values only at $\tau = 0$ and multiples of T for a finite data length $(0 \leq \tau < N)$. Using (5.36) into (5.35), yields

$$r_c(\tau) = \sum_{j=1}^{Q} b_j \sum_{k=0}^{\mu-1} \sum_{l=\tau}^{\infty} h(l-\tau) \left[\frac{(\mu - k)}{N} \delta(l - j - kT) \right] \tag{5.37}$$

Finally, $r_c(\tau)$ can be expressed as

$$r_c(\tau) = \sum_{j=1}^{Q} b_j \sum_{k=0}^{\mu-1} \frac{(\mu - k)}{N} h(kT + j - \tau) \tag{5.38}$$

Clearly, $r_c(\tau)$ depends on the characteristics of the system impulse response $h(n)$. It can be observed from (5.4) that $h(n)$ is a decaying function as $|p| < 1$. In most of the cases, $h(n)$ decays significantly within a period, and thus values of $h(n)$ for $n > T$ can be neglected. Note that in (5.38), except for the case $k = 0$, the system impulse response appears in the form of $h(kT + j - \tau)$ which is a time reversed sequence with respect to $h(\tau)$. Thus, for $k > 0$ in (5.38), a first few lags of $r_c(\tau)$ are contributed by the inner sum due to the presence of some tailing values of the shifted sequences like $h(T + j - \tau), h(2T + j - \tau), \ldots$. The trailing values of these impulse responses residing beyond the first period are generally very small in comparison to the leading values. On the other hand, for $k = 0$, $r_c(\tau)$ exhibits significant values for $0 \leq \tau \leq Q$ and vanishes for any other values of τ because of the causality of the system impulse response. In comparison to these values of $r_c(\tau)$ for $k = 0$, the tailing values, as mentioned before for $k > 0$, can be neglected. When $h(\tau)$ decays sufficiently at a distance L_h from the origin with $L_h < T$, in (5.27), the effect of $r_c(\tau)$ on $r_x(\tau)$ within a period is pronounced at the beginning for $\tau = 0, 1, \ldots Q$ and at the end region for $\tau > T - L_h$. Hence, in the region $Q < \tau < T - L_h$, the effect of $r_c(\tau)$ on $r_x(\tau)$ can be neglected. The above discussion shows that in the case of impulse train excitation,

(5.27) can be simplified as

$$r_x(\tau) = -\sum_{i=1}^{P} a_i r_x(\tau - i), \; Q < \tau < T - L_h \tag{5.39}$$

which can be regarded as the modified version of the ACF obtained in the white noise excitation case as seen from (5.10). As a result, the method proposed in the previous subsection can be readily used to estimate the AR parameters for the impulse train excitation case as long as the value of S is properly chosen such that $P \leq S < T - L_h$ due to the restricted range of $r_x(\tau)$ in (5.39).

The AR parameter estimation scheme proposed in this section offers an advantage of utilizing low-order lags of the ACF which contains more information than the high-order lags. The novelty of the proposed scheme lies in its ability to extract simultaneously the AR parameters and the noise variance from a finite number of observations for white noise or periodic impulse-train excitations. Since, the amount of noise present at the zero lag is treated as an unknown variable instead of depending on noise-subtraction schemes which are very sensitive to the level of noise, the proposed technique can be expected to provide a more accurate AR parameter estimation even at a very low SNR.

5.4 Proposed Method of MA Parameter Estimation

5.4.1 White Noise Excitation

It has been explained in Section 5.2.3 that the MA parameters of the ARMA system can be estimated from the residual signal which is obtained by filtering the observed data via the estimated AR polynomial. In the presence of noise, if the residual signal is directly used without noise-compensation, the estimation performance will

be severely degraded. In what follows, we develop an MA parameter estimation scheme, for which a correlation-domain noise reduction technique is first proposed. The ACF of the noisy residual signal $f_y(n)$ described in (5.14) can be obtained as

$$\begin{aligned} r_{f_y}(\tau) &= E\left[f_y(n)f_y(n+\tau)\right] \\ &= r_{f_x}(\tau) + r_{f_w}(\tau) + r_{f_xf_w}(\tau) + r_{f_wf_x}(\tau) \end{aligned} \tag{5.40}$$

where $r_{f_x}(\tau)$ and $r_{f_w}(\tau)$ are, the ACFs of $f_x(n)$ and $f_w(n)$, respectively, and the last two terms $r_{f_xf_w}(\tau)$ and $r_{f_wf_x}(\tau)$ represent the cross-correlation between $f_x(n)$ and $f_w(n)$. One can express the cross-correlation terms as

$$\begin{aligned} & r_{f_xf_w}(\tau) + r_{f_wf_x}(\tau) = r_{uv}(\tau) + r_{vu}(\tau) \\ & + \sum_{k=1}^{P} \hat{a}_k\{r_{uv}(\tau-k) + r_{vu}(\tau+k)\} \\ & + \sum_{k=1}^{Q} \hat{b}_k\{r_{uv}(\tau+k) + r_{vu}(\tau-k)\} \\ & + \sum_{k=1}^{P}\sum_{l=1}^{Q} \hat{a}_k\hat{b}_l\{r_{uv}(\tau-k+l) + r_{vu}(\tau+k-l)\} \end{aligned} \tag{5.41}$$

Clearly, the above terms are expressed in terms of the shifted cross-correlations between the excitation $u(n)$ and the additive noise $v(n)$. Due to the fact that both $u(n)$ and $v(n)$ are white noise, their cross-correlation values at all lags can be neglected. Thus neglecting the effect of cross-correlation terms on $r_{f_x}(\tau)$, (5.40) can be rewritten as

$$r_{f_y}(\tau) = r_{f_x}(\tau) + r_{f_w}(\tau) \tag{5.42}$$

Now we will investigate each of the terms $r_{f_x}(\tau)$ and $r_{f_w}(\tau)$. From (5.14), $f_x(n)$ can be written as

$$f_x(n) = u(n) + \sum_{l=1}^{Q} \hat{b}_l u(n-l) \tag{5.43}$$

It indicates that $f_x(n)$ can be treated as the output of the MA(Q) system $\hat{B}(z)$ excited by a white noise input $u(n)$ with variance σ_u^2. Thus, $r_{f_x}(\tau)$ can be expressed in terms of the MA parameters $\hat{b}_l$ as

$$r_{f_x}(\tau) = \begin{cases} \sigma_u^2 \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{5.44}$$

Similarly, from (5.14), $f_w(n)$ can be written as

$$f_w(n) = v(n) + \sum_{k=1}^{P} \hat{a}_k v(n-k) \tag{5.45}$$

and $r_{f_w}(\tau)$ can be expressed in terms of the MA parameters $\hat{a}_l$ as

$$r_{f_w}(\tau) = \begin{cases} \sigma_v^2 \sum_{l=0}^{P-\tau} \hat{a}_l \hat{a}_{l+\tau} & |\tau| \leq P \\ 0, & \text{otherwise} \end{cases} \tag{5.46}$$

Thus, using (5.44) and (5.46) into (5.42), $r_{f_y}(\tau)$ can be expressed as

$$r_{f_y}(\tau) = \begin{cases} \sigma_u^2 \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau} + \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, |\tau| \leq Q \\ \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, \ Q < |\tau| \leq P \\ 0, \ |\tau| > P \end{cases} \tag{5.47}$$

In order to estimate the MA parameters, we need to extract an estimate of $r_{f_x}(\tau)$ from $r_{f_y}(\tau)$. It can be observed from (5.47) that, as the estimates of noise variance σ_v^2 and AR parameters $\{a_i\}$ are already obtained (in the previous section), a noise-compensated $r_{f_y}(\tau)$ or an estimate of $r_{f_x}(\tau)$, denoted as $\hat{r}_{f_x}(\tau)$, can be obtained using (5.47) for $|\tau| \leq Q$ as

$$\hat{r}_{f_x}(\tau) = \begin{cases} \hat{r}_{f_y}(\tau) - \hat{\sigma}_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{5.48}$$

As seen from (5.44), the MA parameters $\{b_l\}$ can then be estimated from $\hat{r}_{f_x}(\tau)$, and in this case the estimation of the MA parameters is a nonlinear problem. This problem can be solved by factorizing the polynomial which is constructed using the estimates obtained in (5.48) and given by

$$\hat{S}_{f_x}(z) = \sum_{\tau=-Q}^{\tau=Q} \hat{r}_{f_x}(\tau) z^{-\tau} = \hat{\sigma}_u^2 \hat{B}(z)\hat{B}(z^{-1}) \tag{5.49}$$

In order to compute $r_{f_y}(\tau)$, we employ the scheme proposed in [101], where P_a additional lags after $|\tau| = P$ have been used. From a given ACF, the extracted MA parameters can be made unique only by restricting the process to be causal and/or invertible. It is well-known that the minimum phase system guarantees a causal inverse process. Since the given system is assumed to be minimum-phase, only a valid MA(Q) correlation sequence $\hat{r}_{f_x}(\tau)$ (i.e., the correlation sequence that gives exactly Q zeros inside the unit circle) is used. In order to obtain the MA parameters, the spectral factorization is performed if the estimated ACF sequence belongs to a set of valid MA(Q) correlation sequence. The validity can be checked by using the polynomial-rooting described in [33]. It is clear from (5.48) that the estimation accuracy of the noise variance σ_v^2 would affect the validity of the estimated correlation sequence $\hat{r}_{f_x}(\tau)$. For a better estimation accuracy, a small neighborhood of $\hat{\sigma}_v^2$ is searched to find a valid correlation sequence. Alternatively, validation schemes such as the over-parameterized algorithm of [101] can be used to obtain a valid correlation sequence.

5.4.2 Impulse-train Excitation

In the case of impulse-train excitation $u_i(n)$ with a period T, the noisy residual signal as given by (5.14) can be rewritten as

$$f_{y_i}(n) = \hat{B}(z)u_i(n) + \hat{A}(z)v(n) = f_{x_i}(n) + f_w(n) \tag{5.50}$$

where $f_{x_i}(n)$, the noise-free residual signal can be expressed as

$$\begin{aligned} f_{x_i}(n) &= \sum_{l=0}^{Q} \hat{b}_l u_i(n-l), \; b_0 = 1 \\ &= \sum_{l=0}^{Q} \sum_{k=0}^{\lambda-1} \hat{b}_l \delta(n - kT - l) \end{aligned} \tag{5.51}$$

and $f_w(n)$ is given by (5.45). It can be observed from (5.50) that $f_{x_i}(n)$ is the output of an MA(Q) system $\hat{B}(z)$ excited by $u_i(n)$, and $f_w(n)$ is an MA(P) sequence excited by $v(n)$. Since the impulse response of an MA(Q) system, say $h_Q(n)$, vanishes beyond $n = Q$ samples, $f_{x_i}(n)$ is a periodic repetition of $h_Q(n)$ without overlaps when $T > Q$. Using (5.6) and (5.51), an expression for $r_{f_{x_i}}(\tau)$ can be derived as

$$r_{f_{x_i}}(\tau) = \sum_{k=-(\lambda-1)}^{(\lambda-1)} (\lambda - |k|)\, r_{h_Q}(\tau - kT) \tag{5.52}$$

where

$$r_{h_Q}(\tau) = \begin{cases} \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{5.53}$$

Recall that λ is the total number of impulses in the impulse-train excitation defined in (5.23). It can be observed from (5.52) and (5.53) that $r_{f_{x_i}}(\tau)$ is periodic with period T, and therefore, there will be no aliasing in the correlation domain if $T > 2Q + 1$. In the region $|\tau| < T - Q$, $r_{f_{x_i}}(\tau)$ reduces to

$$r_{f_{x_{iT}}}(\tau) = \begin{cases} \lambda \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau}, & |\tau| \leq Q \\ 0, & Q < |\tau| < T - Q \end{cases} \tag{5.54}$$

The autocorrelation of the noisy residual signal $f_{y_i}(n)$ can be expressed as

$$r_{f_{y_i}}(\tau) = r_{f_{x_i}}(\tau) + r_{f_w}(\tau) \tag{5.55}$$

where $r_{f_w}(\tau)$ is given by (5.46). Note that, as described in the white noise excitation case (see (5.41)), the cross-correlation terms between $f_{x_i}(n)$ and $f_w(n)$ in (5.55) has been neglected, since they consist of cross-correlations between $u_i(n)$ and $v(n)$, which can be considered to be uncorrelated. As seen from (5.55), the component $r_{f_{x_i}}(\tau)$ is a periodically repeated scaled version of $r_{h_Q}(\tau)$ (see (5.52)), and the component $r_{f_w}(\tau)$ vanishes beyond $\tau = P$ samples. Using (5.46) and (5.54), for $T > P + Q + 1$, (5.55) can be rewritten as

$$r_{f_{y_i}}(\tau) = \begin{cases} \lambda \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau} + \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, & |\tau| \leq Q \\ \sigma_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, & Q < |\tau| \leq P \\ r_{f_{x_i}}(\tau), & |\tau| > P \end{cases} \tag{5.56}$$

Note that the above form of $r_{f_{y_i}}(\tau)$ is similar to that of $r_{f_y}(\tau)$ given by (5.47). As the effect of $r_{f_w}(\tau)$ vanishes after $\tau = Q$, only $r_{f_x}(\tau)$ exists after $\tau = P$. From (5.52) and (5.56) it can be observed that a non-zero value of $r_{f_{y_i}}(\tau)$ beyond $\tau = P$ appears only after $\tau = T - Q$ due to $r_{f_x}(\tau)$. The number of additional lags (P_a) to be used in the computation of $r_{f_{y_i}}(\tau)$ is restricted to $P_a < T - (P + Q + 1)$. In practice, $T \gg P$ and thus, the condition $T > P + Q + 1$ will be automatically satisfied. Thus, an estimate of $r_{f_{x_i}}(\tau)$ can be obtained as

$$\hat{r}_{f_{x_i}}(\tau) = \begin{cases} \hat{r}_{f_{y_i}}(\tau) - \tilde{\hat{\sigma}}_v^2 \sum_{k=0}^{P-\tau} \hat{a}_k \hat{a}_{k+\tau}, & |\tau| \leq Q \\ 0, & \text{otherwise} \end{cases} \tag{5.57}$$

in which $\hat{r}_{f_{x_i}}(\tau)$ has been set to zero for $\tau > Q$. In the original expression for $r_{f_{x_i}}(\tau)$

as given by (5.52), when $r_{f_{x_i}}(\tau)$ is set to zero beyond $\tau = Q$, it can be written as

$$r_{f_{x_i}}(\tau) = \lambda \sum_{l=0}^{Q-\tau} \hat{b}_l \hat{b}_{l+\tau} = \lambda r_{h_Q}(\tau), |\tau| \leq Q \tag{5.58}$$

It is clear from (5.56), (5.57) and (5.58) that $\hat{r}_{f_{x_i}}(\tau)$ is the estimate of $r_{f_{x_i}}(\tau)$, which is a scaled version of $r_{h_Q}(\tau)$ given by (5.53). Using the estimates $\hat{r}_{f_{x_i}}(\tau)$, similar to (5.49) that is obtained for the case of white noise excitation, a polynomial can be constructed as

$$\hat{S}_{f_{x_i}}(z) = \sum_{\tau=-Q}^{\tau=Q} \hat{r}_{f_{x_i}}(\tau) z^{-\tau} = \lambda \hat{B}(z) \hat{B}(z^{-1}) \tag{5.59}$$

Thus, the MA parameters can be estimated using spectral factorization once the validity check for $\hat{r}_{f_{x_i}}(\tau)$ has been carried out.

The main steps of the complete algorithm are summarized as Algorithm I.

Algorithm I: The Proposed NCCD Method

1. Compute the autocorrelation $r_y(\tau)$ from noisy observation $y(n)$ using (5.6).

2. Determine the AR parameters and the noise variance using the technique (Section 5.3) comprising the following steps:

   i. Compute $\mathbf{\Theta}_0$, $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ which are required in the quadratic eigenvalue problem (5.21).

   ii. Calculate real square matrices $\mathbf{E}$ and $\mathbf{F}$ using (5.21).

   iii. Solve (5.21) using the QZ algorithm [132].

   iv. From the solution of (5.21), find the minimum eigenvalue and use the corresponding eigenvector in (5.22) to estimate the AR parameters.

3. Determine the residual signal $f_y(n)$ by filtering $y(n)$ using the estimated AR polynomial obtained in Step 2 and estimate its autocorrelation $r_{f_y}(\tau)$.

4. Obtain the noise-compensated ACF $\hat{r}_{f_x}(\tau)$ of the residual signal using (5.48).

5. Estimate the MA parameters from $\hat{r}_{f_x}(\tau)$ using spectral factorization based on (5.49).

5.5 Simulation Results

A number of simulations is carried out for the identification of ARMA systems under noisy conditions, and results along with some comparative analysis are investigated in this section. We consider synthetic signals generated from ARMA systems of different orders as well as natural speech signals corrupted by additive noise. Next, the performance in terms of the accuracy and consistency of the estimated parameters of the proposed method is obtained and compared with that of the conditional maximum-likelihood (ML) method [34], also referred to as the prediction error method, the ARMA cepstrum recursion (ACR) method [45], and the reduced statistics (RS) method [33], [133].

5.5.1 Results on Synthetic ARMA Systems

(a) White Noise Excitation: A noisy ARMA signal is generated according to (5.1) and (5.11) with $N = 4,000$ and $\sigma_u^2 = 1$, where the variance of the white Gaussian noise can be appropriately set according to the desired SNR defined as

$$\text{SNR} = 10 \log_{10} \left[\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} v(n)^2} \right] \quad \text{dB} \tag{5.60}$$

The number of equations in (5.19) is governed by S, and we have chosen $S = 2P$ in order to avoid very high order lags of the ACF in the computation of the AR parameters. The QZ algorithm which is required to solve (5.21) is implemented through MATLAB built-in command '*eig*'. In the estimation of the MA parameters,

$P_a = 60$ additional lags of $\hat{r}_{f_y}(\tau)$ is used. In the case, $\widetilde{\hat{\sigma}}_v^2$ cannot produce a valid MA(Q) correlation sequence, a neighborhood search, say in the region of 5% of $\widetilde{\hat{\sigma}}_v^2$ with a resolution of 0.01 is performed. As a matter of fact, in our extensive simulations, it has been observed that in most of the cases, less than 10 searches are sufficient to obtain a valid correlation sequence.

Each experiment contains $N_T = 100$ independent trials. We have conducted the experiments for the noisy cases, where 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) estimation mean, (2) the standard deviation from the mean (SDM), (3) the standard deviation from the given value (or the true value) (SDT), and (4) the average sum-squared error (ASSE) given by

$$\text{ASSE} = \frac{1}{N_T(P+Q)} \sum_{m=1}^{N_T} (\varepsilon_a + \varepsilon_b) \tag{5.61}$$

where $\varepsilon_a = \sum_{k=1}^{P} [\hat{a}_k(m) - a_k]^2$, $\varepsilon_b = \sum_{j=1}^{Q} [\hat{b}_j(m) - b_j]^2$, and $\hat{a}_k(m)$ and $\hat{b}_j(m)$ are the estimated parameters at the mth trial, and a_k and b_j the true values of the parameters. Different ARMA systems are investigated in order to cover a wide range of possible combinations of pole-zero locations as well as types (i.e., real or complex conjugate). Moreover, since the real and complex types of zeros or poles exhibit quite different behaviors, in our experiments various combinations of real and complex poles and zeros are considered to show the capability of the proposed algorithm in dealing with real life situations.

In Table 5.1, the estimation results for the ARMA(3, 2) system in terms of the four performance measurement criteria are shown at SNR= 0 dB. In this table, the last row gives the true parameter values. The estimated values of the corresponding parameters obtained from the proposed and the three other methods are given in

Table 5.1: Estimated parameters along with standard deviations (SDM and SDT) for white noise excited ARMA(3, 2) system at SNR = 0 dB

Methods	Estimated parameters					ASSE
	a_1	a_2	a_3	b_1	b_2	(dB)
Proposed	−2.5938	2.5592	−0.9689	−1.6379	0.8394	−18.27
	(±0.0182)	(±0.0321)	(±0.0217)	(±0.0562)	(±0.0639)	
	(±0.0187)	(±0.0384)	(±0.0241)	(±0.0584)	(±0.0657)	
ML	−1.7840	1.3414	−0.2546	−1.3236	0.6665	−2.38
	(±0.3452)	(±0.2662)	(±0.0487)	(±0.3447)	(±0.1970)	
	(±0.7509)	(±1.2100)	(±0.6924)	(±0.4362)	(±0.2241)	
RS	−1.8944	1.5783	−0.4057	−1.2960	0.5777	−5.098
	(±0.0917)	(±0.1516)	(±0.0625)	(±0.0760)	(±0.0767)	
	(±0.7456)	(±0.9012)	(±0.5427)	(±0.4287)	(±0.2801)	
ACR	−2.2734	2.1750	−0.6652	−1.2052	0.8521	−10.73
	(±0.1164)	(±0.1542)	(±0.1798)	(±0.0492)	(±0.0512)	
	(±0.3259)	(±0.4795)	(±0.2968)	(±0.4882)	(±0.0554)	
True	−2.5712	2.5218	−0.9460	−1.6909	0.8100	

the preceding four rows. The corresponding values for the SDM and SDT are shown within the parentheses below the each estimated parameter value. The last column of each table provides the ASSE in dB. It is to be mentioned that to implement the ML method the '*armax*' command from the MATLAB System Identification Toolbox is used [34]. It is seen from the table that the proposed method exhibits superior performance with respect to all the four performance indices. Very small values of SDM and SDT obtained from the proposed technique indicate a high degree of estimation consistency and accuracy.

Fig. 5.2 shows the ASSE values as a function of SNR levels for the four methods for the ARMA(4, 3) system with parameters $a_i = 1, -2.7303, 3.7012, -2.5906, 0.9036$ and $b_j = 1, -2.0739, 1.7219, -0.6167$. It is observed from Fig. 5.2 that the estimation error occurred in case of the proposed method is significantly lower throughout the whole range of SNR.

Fig. 5.3 shows the superimposed plots of the estimated poles and zeros from 20 realizations obtained by using the different methods at SNR = 0 dB for the ARMA(5, 4)

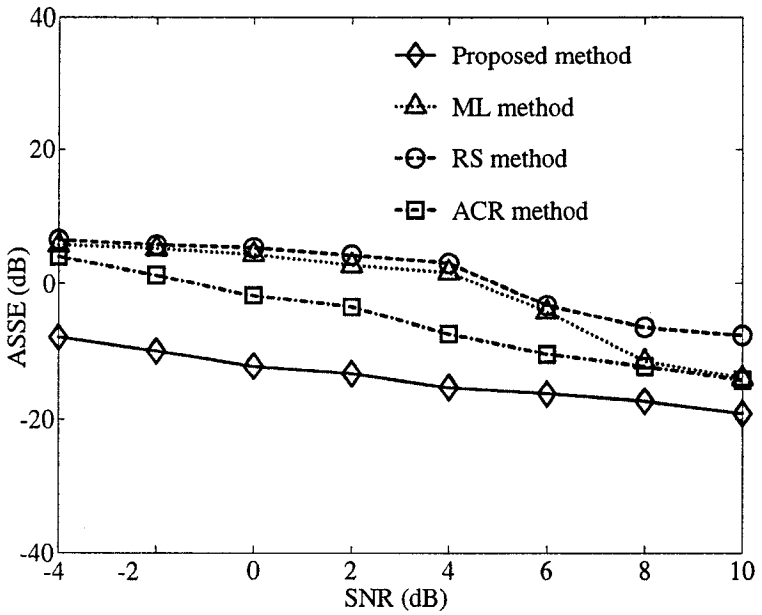


Figure 5.2: Effect of SNR on the estimation accuracy for the white noise excited ARMA(4, 3) system.

system whose parameters are $\{a_k\} = \{1, -2.0825, 2.267, -2.1997, 1.8563, -0.811\}$ and $\{b_j\} = \{1, -1.6379, 1.5279, -1.2989, 0.6281\}$. For the purpose of comparison, the true poles and zeros are also plotted. A major advantage of the proposed method as clearly observed from the figure is that the estimated pole-zero values are much less scattered around the true values indicating a very high estimation accuracy in comparison to that achieved by the other methods.

(b) Impulse-Train Excitation: We have also considered the problem of ARMA SI with the periodic impulse-train excitations of different periods for various levels of noise. An impulse-train is generated using (5.23) with a known value of T. A noisy ARMA signal is generated according to (5.1) and (5.11) with $N = 4,000$. The simulations are run over $N_T = 100$ independent trials and the results averaged.

The ASSE resulting from using various methods under the impulse-train excitation

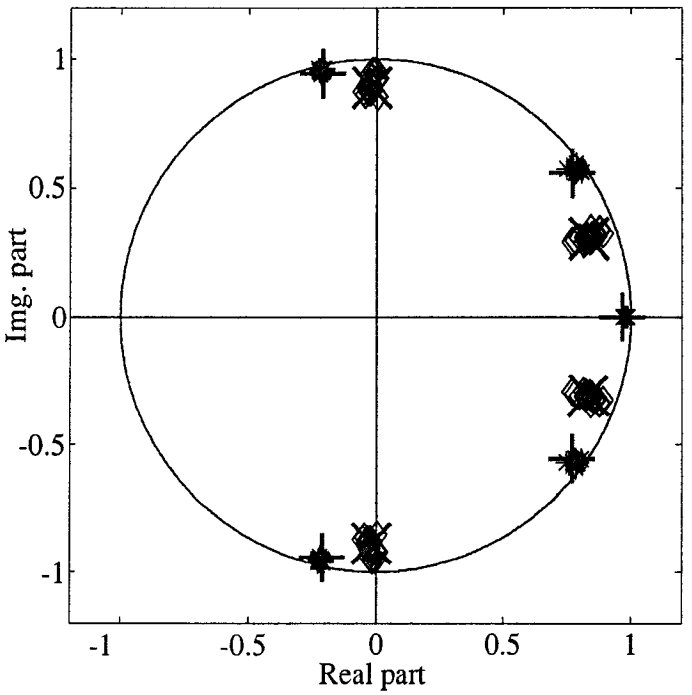


Figure 5.3: Superimposed pole-zero plot of ARMA(5, 4) system at SNR = 0 dB obtained by the proposed method. + : true poles, × : true zeros, ∗ : estimated poles, ◇ : estimated zeros.

with $T = 70$ for the estimation of the same ARMA(3, 2) system as the one considered for the white noise excitation is shown in Fig. 5.4. As seen, the proposed method provides a significantly better performance even at a very low SNR, whereas the performance of other methods deteriorates at SNR level below 10 dB.

5.5.2 An Application for Vocal-tract System Identification

Proposed identification scheme has been tested for the purpose of pole-zero estimation of the vocal-tract system with natural speech signals in the presence of noise. In Fig. 5.5(a), average estimated values of poles and zeros obtained from 20 independent trials are shown for a speech phoneme /m/ of the word '*him*' uttered by a female speaker and taken from TIMIT standard database. Here the estimation is carried

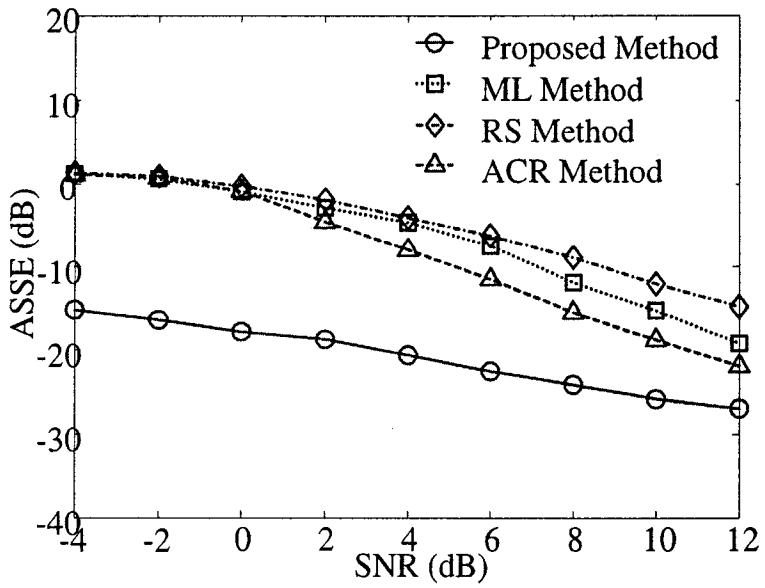
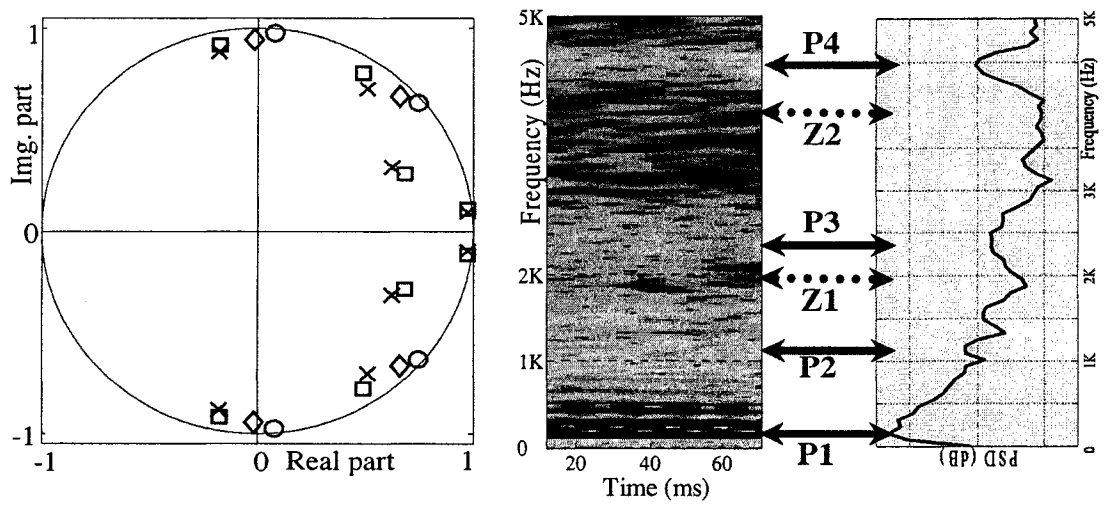


Figure 5.4: Effect of SNR on the estimation accuracy for the impulse-train excited ARMA(4, 3) system.

out in the presence of AWGN at SNR = 0 dB considering an ARMA(8, 4) system. In this implementation, hamming window is employed and an FFT pre-filtering is performed. No pre-emphasis is done. In Fig. 5.5(a), for reference values, pole-zero locations estimated by the ML method at a noise-free condition are plotted. In Fig. 5.5(b) the spectrogram and the PSD of the clean speech are presented to indicate the pole-zero (P_i, Z_i) locations. This figure clearly exhibits the estimation accuracy of the proposed method in the case of natural signals.

5.6 Conclusion

In this chapter, a new scheme for the identification of the ARMA systems under noisy conditions has been presented. It has been shown that the proposed identification scheme has the ability to estimate the ARMA system parameters with sufficient



(a) (b)

Figure 5.5: For natural sound /m/, (a) pole-zero estimates of the Proposed method at SNR = 5 dB are overlaid on that of the ML method at clean condition (Poles: $\times$: Clean, $\square$: proposed; Zeros: O: clean, $\diamond$:proposed). (b) Spectrogram and PSD of noise-free speech.

accuracy and consistency using only the noise-corrupted observations, without the need for accessing the input or noise-free output data. In order to overcome the limitations of the conventional Yule-Walker methods, by separating the noise variance from the zero lag of the ACF, a quadratic eigenvalue problem is formulated and solved for joint estimation of the AR parameters as well as the observation noise variance. The estimation accuracy of the proposed AR estimation technique is increased as a result of utilizing all possible low-order lags of the ACF which contain more system information than the high-order lags. A significant feature of the new technique is that it has been presented in a unified manner for both white noise and periodic impulse-train excitations. Unlike conventional methods, noise-compensation is performed for the identification of both the AR and MA parts. For the AR part, since the AR

parameters have been estimated together with the noise variance, no separate noise compensation has been performed. For the MA part, an explicit noise-compensation algorithm has been proposed based on the estimated noise variance and the AR parameters, which efficiently reduces the effect of noise on the correlation function of the residual signal.

In the proposed method, the system order is assumed to be known. As mentioned in Chapter 2, in the case of real-life data (i.e., the data in the presence of noise) a bad choice of the model order would affect the estimation accuracy of the parameters for all the methods. Unlike model-fitting based methods described in Chapters 2, 3, and 4, the proposed method does not involve search operations for the parameter estimation. Thus, the computational complexity of the proposed method, in comparison to the model-fitting based methods, is low and the computational time required is also found quite low for practical applications where the objective is to achieve an accurate estimation of the system parameters at very low levels of SNR.

Extensive simulations have been carried out to demonstrate the performance of the proposed technique. It has been shown that the new scheme outperforms the other existing methods considered for comparison. As a practical application of the proposed technique, the identification of human vocal-tract system in the presence of noise has been attempted, yielding quite accurate system parameters or poles.

Some of the distinctive features of the proposed method reinforced by the experimental results can be summarized as follows.

1. The proposed method utilizes all possible low-order lags of the ACF, which contain more system information than the high-order lags, in order to obtain a better AR parameter estimation accuracy.

2. An important advantage of the proposed method is that it can estimate the AR parameters together with the unknown additive noise variance.

3. In the proposed method, the input excitation power need not be assumed to be known.

4. Because of the noise-compensation from the ACF of the residual signal, a better accuracy in the MA parameter estimation is obtained.

5. It is capable of handling the problem of ARMA system identification under severe noisy conditions for both white noise and impulse-train input excitations. This feature makes the method readily applicable to speech signals.

Chapter 6

Formant Frequency Estimation of Speech Signals

6.1 Introduction

Formant is one of the most informative speech features used in interpretation of the mechanism of human speech production. Formant frequency estimation has found a wide range of applications in speech analysis/synthesis, coding, and recognition [82]. Most of the formant frequency estimation methods so far reported, are capable of handling only the noise-free environments [84], [85], [87]–[92]. In view of real-life applications, formant frequency estimation from noise-corrupted speech is an essential but difficult task. Only a few formant estimation methods are available in literature which deal with noisy environments [93], [94]. However, these methods give very high errors in the estimation of higher formants.(i.e. other than the first formant).

In this chapter, new methods for the estimation of formant frequencies from noise-corrupted speech signals are presented. The main target is to develop formant estimation schemes based on our new system identification methods which can efficiently tackle the adverse effect of observation noise and provide an accurate estimate of formant frequencies of speech signals. Since vocal-tract resonances correspond to

formants, the task of formant frequency estimation can be treated mainly as a vocal-tract system identification. The formant estimation schemes proposed in this thesis can be classified into three groups depending on the main principle of formant frequency estimation: (1) Correlation based methods [134]–[138], (2) Ramp cepstrum based methods [139]–[142], and (3) Ramp cosine cepstrum based methods [143]–[145].

Among different formant estimation techniques, correlation based methods, such as different variants of linear predictive coding (LPC) methods, are most commonly used [7], [84], [85]. However, under a noisy condition, the estimation performance of the LPC based formant estimation methods deteriorates significantly. Recently in [134], [138], correlation model based formant frequency estimation methods has been proposed in order to tackle very severe noisy environments. In these methods the overall vocal-tract system is considered as an ARMA model and based on the correlation model proposed in Chapter 2, a correlation model-fitting approach is proposed in conjunction with an adaptive residue based least square (RBLS) optimization algorithm. In [135], a correlation domain method is developed wherein formant frequencies are estimated from noise-compensated speech signals using a modified form of least-squares Yule-Walker (LSYW) method along with an effective formant selection criterion. In [137], a once-repeated autocorrelation function (ORACF) of the observed noisy signal is employed in a modified form of least-squares Yule-Walker (LSYW) equations which provides a better formant frequency estimates. Moreover, instead of directly using the estimated poles for extracting the desired formant frequencies, a frequency-domain peak-picking algorithm within a certain band is introduced which enables the proposed scheme to avoid estimation errors that may occur in the case of weak formants.

Cepstral domain formant estimation methods available in literature mostly deal with noise-free environments [87], [86]. In [139], a new technique for the formant frequency estimation from noise-corrupted speech data is presented based on the ramp-cepstrum model for the OSACF of AR signals proposed in Chapter 3. In this method, the voiced speech signal is considered as the the output of the AR model with a periodic impulse-train excitation. A ramp-cepstrum model-fitting based approach is proposed where the adaptive RBLS optimization algorithm is used in order to obtain an accurate estimate of the ramp-cepstrum model parameters which gives the desired formant frequencies in the presence of significant noise. Recently, in [140], a ramp cepstrum model of a once-repeated autocorrelation function (ORACF) of the voiced speech signal in terms of formant parameters is proposed and ramp-cepstrum model-fitting is used to obtain formant frequencies from noisy observations.

In Chapter 4, we proposed a ramp cosine cepstrum (RCC) model for the OSACF of AR signals. In order to obtain an accurate estimate of formant frequencies under a heavy noisy condition, we now develop a model for the RCC of the ORACF of speech signals [143], [144]. Since the proposed RCC model provides a direct relationship with formant frequencies, the RBLS optimization algorithm is used to solve the model-fitting problem of formant estimation in the presence of severe noise. The DCT-IDCT is then employed for the purpose of RCC implementation.

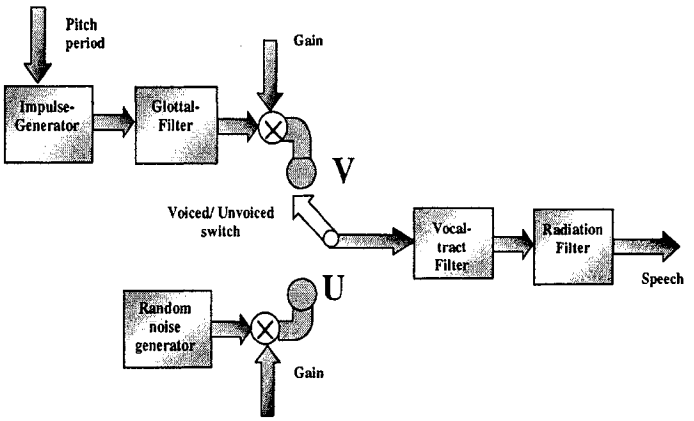
The rest of the chapter is organized as follows. In Section 6.2, first, a background of formant estimation is presented. Then, a framework that is used to develop the new formant estimation methods is proposed in 6.3. In Section 6.4, the ARMA correlation model-fitting based formant estimation method is presented. In Section 6.5, a ramp-cepstrum model-fitting based formant estimation method is described.

In section 6.6, a ramp cosine cepstrum model for once-repeated ACF of the observed speech is first proposed, based on which a formant estimation method based on the RCC model-fitting approach is then developed. Finally, some concluding remarks are given in 6.7.

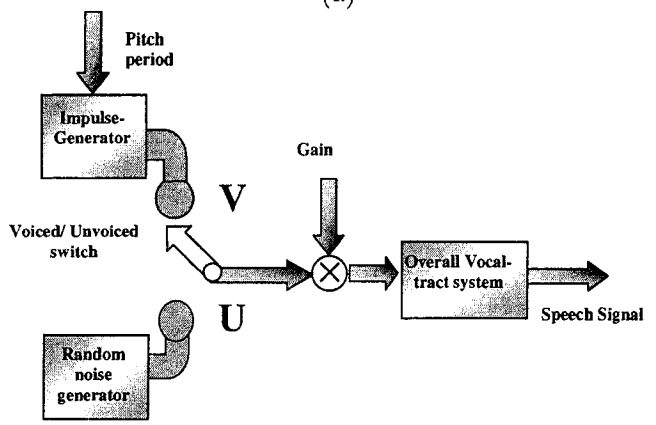
6.2 Speech Production System and Formant Estimation

Formant is one of the most useful speech features used in interpretation of the mechanism of human speech production. It is also well related to the articulatory activity and the perception of speech signals.

Fig. 6.1(a) represents a human speech production system which is basically composed of the excitation model, the vocal-tract filter, and the radiation filter. For an unvoiced (U) speech, the excitation can be considered as a flat spectrum noise source normally modeled by a random noise generator. For a voiced (V) speech, the excitation model contains an impulse generator producing the impulse-train at the pitch period and a glottal pulse shaping filter driven by the impulse-train to generate the excitation signal. In a simplified or widely used working model of speech production system, the glottal pulse shaping filter, the vocal-tract filter, and the radiation filter are combined to form a single overall vocal-tract system as shown in Fig. 6.1(b) [7], [8]. Then, the overall system is driven by the periodic impulse train generator resulting in voiced speech. In this model, obviously, the glottal pulse shaping filter is not required for the unvoiced sound and the excitation source is the random noise generator generally modeled as a white Gaussian noise. In general, the overall vocal-tract system can be treated as an acoustic resonator, where the resonance frequencies are given by system poles. Formants of speech signals are usually attributed to the resonance



(a)



(b)

Figure 6.1: Discrete-time speech production model. (a) Detailed model and (b) working model.

frequencies of the vocal tract. The resonant frequency of each significant pole of the vocal-tract system is a formant candidate. For the formant frequency estimation it is sufficient to restrict the analysis to the voiced speech where the excitation for the VT can be modeled as the output of a glottal filter whose input is the periodic impulse-train.

The overall vocal-tract filter of a human speech production system can be represented by an AR (all-pole) or ARMA (pole-zero) model depending on the character-

istics of a particular phoneme. For example, vowel sounds are well represented by AR model, whereas some nasal sounds like /m/ or /n/ require ARMA models [7]. The transfer function of a VT system represented by an AR(P) model is given by

$$H(z) = \frac{G}{A(z)} = \frac{G}{\prod_{k=1}^{P}(1 - p_k z^{-1})} \tag{6.1}$$

and that of the corresponding ARMA(P, Q) system is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{\prod_{j=1}^{Q}(1 - z_j z^{-1})}{\prod_{k=1}^{P}(1 - p_k z^{-1})} = \sum_{k=1}^{P} \frac{\eta_k}{1 - p_k z^{-1}} \tag{6.2}$$

where G is the gain, $A(z) = \sum_{k=1}^{P} a_k z^{-k}$ is the AR polynomial with AR parameters $\{a_i\}$, $B(z) = 1 + \sum_{j=1}^{Q} b_j z^{-j}$ gives the MA part of the ARMA system with MA parameters $\{b_j\}$, p_k's and z_j's denote, respectively, the poles and the zeros, and η_k the partial fraction coefficient corresponding to the kth pole. Here, P and Q are, respectively, AR and MA orders. In the case of ARMA(P, Q) model generally it is assumed that $P > Q$. Note that, for real-life data, such as speech signal, it is sufficient to use an ARMA model with less number of zeros than the number of poles [7]. The number of poles and zeros, which is required to represent different speech phonemes, has been investigated by several researchers [146]. Although the standard techniques for order estimation are available in the literature [98], [99], the method we are going to develop does not require the *a priori* knowledge of orders, since it attempts to estimate only a few formants from certain range of poles. In order to model each formant, a pair of complex conjugate poles is required. Formant frequency (F_k) and bandwidth (B_k) can be computed from the system pole, $p_k = r_k e^{j\omega_k}$ and the sampling

frequency F_s as [93]

$$\begin{aligned} F_k &= \frac{F_s}{2\pi}\omega_k; \\ B_k &= -\frac{F_s}{\pi}\ln(r_k) \end{aligned} \tag{6.3}$$

The pole angle ω_k relates to the resonant frequency and the pole radius r_k is related to the concentration of local energy and the bandwidth of spectral resonance of a formant candidate. Depending on the speaker characteristics and the phonemes, typically first few (three to five) formants are required in practical applications and also these formant locations are specified within a frequency range of $0-5$ kHz [7], [88]. Hence, by identifying the poles of vocal tract system, formant frequencies can be estimated. Estimating formants accurately from natural speech is not so easy because of the variety of speech sounds. Some major problems involved in formant estimation are summarized below:

1. Time-varying characteristics in the glottal open phase of the vocal-tract system.
2. The size, position, and shape of the analysis frame.
3. Low spacing between the adjacent formants, which in the worst case may cause one broad spectral peak instead of two distinct peaks
4. Very high values of the fundamental frequency (or pitch).
5. The presence of observation noise corrupting the speech.

The first four problems can be treated as some inherent problems,which exist no matter whether the formant estimation is performed in the presence of observation noise or not. However, in the proposed formant estimation methods, the effect of these problems can be significantly reduced since our methods do not work directly in the

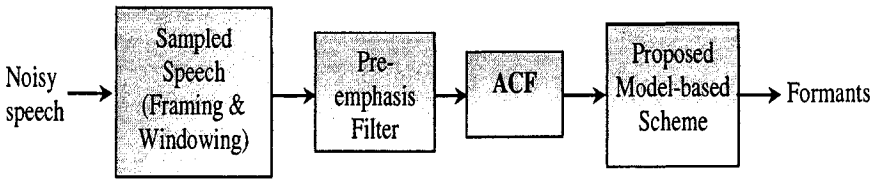


Figure 6.2: Block diagram of the proposed framework for the formant frequency estimation method.

signal domain, rather they work in the correlation domain or in the cepstral domain which is computed from a correlation function. The problem of the observation noise is the most crucial one and the proposed methods are targeted to tackle this problem through new correlation and cepstral domain model fitting approaches to be developed in the next sections.

6.3 Proposed Framework for Formant Estimation Methods

A general block diagram showing some common steps involved in all of our proposed formant estimation methods is shown in Fig. 6.2. First, the observed noisy signal is pre-processed with sampling and windowing. Then the windowed speech frame is pre-emphasized. A pre-filtering for very low frequencies is also performed. Note that all of the proposed formant frequency estimation methods need to perform the autocorrelation of the speech signal $x(n)$ at the first stage which can be computed,in general, as

$$r_x(m) = \frac{1}{N} \sum_{n=0}^{N-1-|m|} x(n)x(n+|m|),\ 0 \leq |m| < N \tag{6.4}$$

where N is the data length [134].

Then, the correlation values are used in different algorithms. Also, an initial estimate of the formant frequencies is required in the adaptive RBLS algorithm to

reduce the search space. In what follows, we are first going to describe the pre-processing and initial frequency estimation steps in detail. In the next sections, we will describe three formant estimation methods based on the proposed models, namely (1) ARMA correlation model, (2) ramp-cepstrum model, and (3) ramp cosine cepstral model.

6.3.1 Preprocessing

In the proposed method, formant estimation is performed on a frame by frame basis. In order to carry out the short-time analysis on the observed speech signals, first, windowing is performed so as to reduce the edge effects at the beginning and the end of the frame. A Hamming window of length N is used with a certain percentage of overlap. For the purpose of analysis, the speech signal is considered to be stationary over the short observation interval. Since voiced speech spectra normally have a roll-off of about -6 dB/Octave, they are tilted into a slightly low-pass form [7]. In order to reduce the natural spectral tilt of the windowed speech signal, a high-pass pre-emphasis filter with the following input-output relation is employed

$$s(n) = s'(n) - \gamma_p s'(n-1) \tag{6.5}$$

where $s(n)$ is the filtered output with respect to the input $s'(n)$ with a pre-emphasis factor $\gamma_p < 1$. The effect of applying pre-emphasis filter can be viewed as introducing an extra zero into the transfer function of the vocal tract filter. Note that the introduction of such a zero neither alters the pole locations in the transfer function plane, nor affects their resonance center frequencies and bandwidths in the associated frequency response. In Figs. 6.3 and 6.4, the effect of windowing and pre-emphasis is demonstrated in the time- and frequency-domain, respectively. Fig. 6.3(a) shows

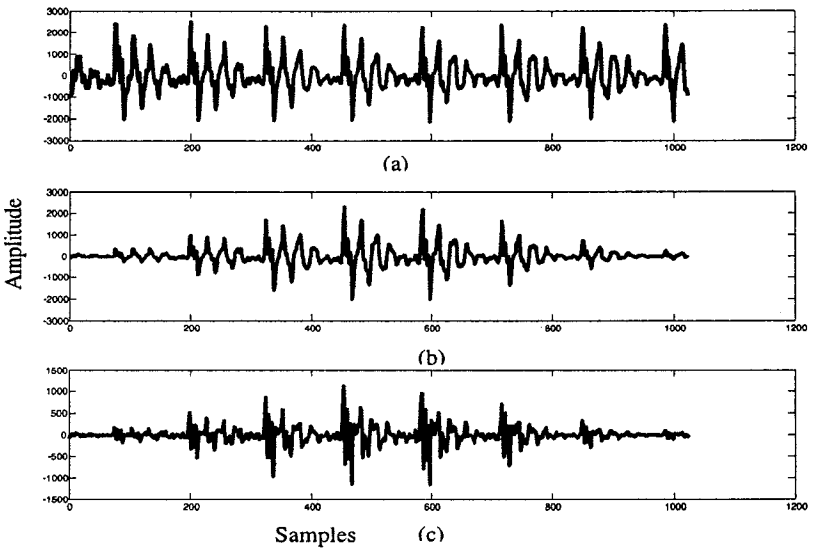


Figure 6.3: Effect of windowing and pre-emphasis on the observed speech signal in time domain. (a) clean speech, (b) windowed speech obtained from (a) using hamming window, (c) pre-emphasized speech obtained from (b).

the clean speech, Fig. 6.3(b) displays the corresponding hamming windowed speech and Fig. 6.3(c) gives the pre-emphasized version of the windowed speech. Fig. 6.4(a) to 6.4(c) demonstrate the power spectral density plots corresponding to Fig. 6.3. In Fig. 6.4, the first three formant frequencies are prominent with a natural spectral tilt. It is clearly observed from Fig. 6.4(c) that the spectral tilt is reduced, and the second and third formant frequencies are significantly lifted due to the pre-emphasis operation.

6.3.2 Pre-filtering and Initial Formant Estimation

In order to obtain an initial estimate of formant frequencies from observed speech signals, generally, a peak-picking operation is performed on the spectral domain. For correlation domain methods, peaks from the smoothed power spectral density plot are computed. In the case of using cepstral domain methods, an initial estimate of

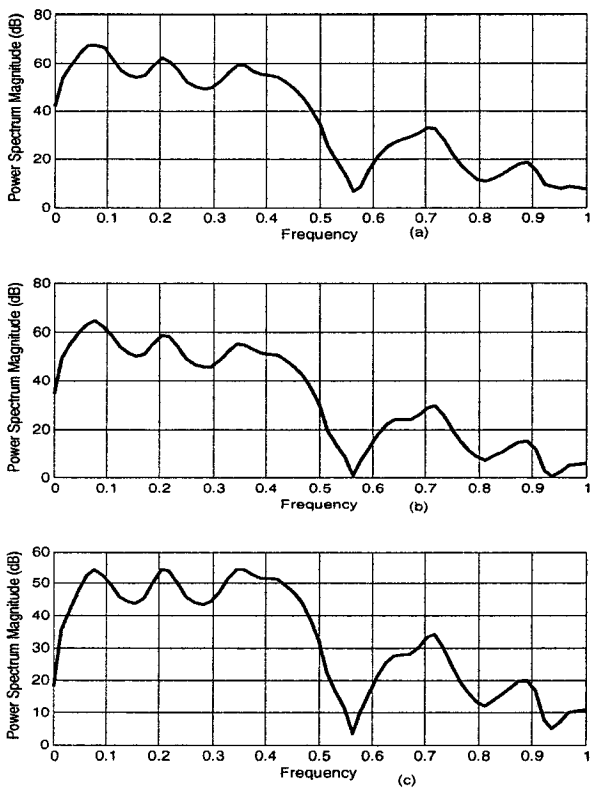


Figure 6.4: Effect of windowing and pre-emphasis on the observed speech signal in frequency domain. (a) clean speech, (b) windowed speech obtained from (a) using hamming window, (c) pre-emphasized speech obtained from (b).

the most dominant formants can be obtained from the peaks of the smoothed DCT magnitude spectrum. These possible candidates of the formant frequencies are used in the least-square optimization process. Note that, it is described later on that in the presence of heavy noise, instead of the ACF of noisy observations, its ORACF offers more noise robustness. In addition to signal pre-emphasis, a pre-filtering is performed to remove the effect of very low frequencies ($<$ 100 Hz) which are not of our interest. In fact, this will also- suppress the effects of the pitch from the

first-formant estimation. Pre-filtering can be either performed in the spectral domain using FFT-IFFT or in the case of cepstral domain using DCT/IDCT. The signal $x(n)$ obtained after the preprocessing and pre-filtering is used for the next step of least-square model-fitting. Next, we will first develop the desired model and then describe the least-square model fitting approach to estimate the formant frequencies.

6.4 ARMA Correlation Model Based Formant Estimation

6.4.1 Proposed Method

In Chapter 2, a correlation model for the output of an ARMA system excited by a periodic impulse-train has been derived. Since, within a short duration of time, the speech can be considered as the output of an ARMA system excited by a periodic impulse-train (with pitch period) for a voiced speech segment, the proposed ARMA correlation (ARMAC) model can be used to extract the poles of the overall vocal tract system. As discussed in the previous section, formant frequencies can be estimated from the poles of the overall vocal-tract system. Hence, the main task is now reduced to the accurate estimation of the ARMAC model parameters giving the system poles. Unlike conventional correlation based methods, a correlation-fitting based approach is proposed where an adaptive RBLS optimization algorithm is introduced in order to obtain an accurate estimation of the model parameters from noise-corrupted output observations even in the presence of significant noise.

During a short duration of time (frame), a given speech signal $x(n)$ is generally assumed to be stationary. Hence, the overall vocal-tract transfer function $H(z)$ given by (6.2) can be modeled with constant coefficients within a short-time window of speech. The input excitation for voiced sounds is modeled as the periodic impulse-

train $\{u_i(n)\}_{n=0}^{N-1}$ with period T as given by

$$u_i(n) = \sum_{m=0}^{\lambda-1} \delta(n - mT), \quad \lambda = \lceil N/T \rceil \tag{6.6}$$

where $\lceil \zeta \rceil$ represents the smallest integer greater than or equal to ζ, and thus, λ is the total number of impulses in the excitation. The ARMA correlation model that we have derived in Chapter 2 is given by

$$r_x(\tau) = \sum_{l=1}^{\theta} r_l^{\tau}[\alpha_l \cos(\omega_l \tau) + \beta_l \sin(\omega_l \tau)], \tau = 0, 1, \ldots, M-1 \tag{6.7}$$

where $\alpha_l = \zeta_l \cos \upsilon_l$, $\beta_l = -\zeta_l \sin \upsilon_l$. As mentioned before, since each formant corresponds to a pair of complex conjugate poles $p_k = r_k e^{\pm j\omega_k}$, for the case of formant analysis it is sufficient to consider only the complex conjugate poles. Thus, θ in (6.7) is the number of pairs of complex conjugate poles. Each of the θ terms in the summation of (6.7), namely $F_l(\tau) = r_l^{\tau}[\alpha_l \cos(\omega_l \tau) + \beta_l \sin(\omega_l \tau)]$, can be estimated sequentially from the M lags of $r_y(\tau)$, which is available from the noisy observations.

Considering the observation noise $v(n)$ as an additive white Gaussian, the noise-corrupted speech signal $y(n)$ is given by

$$y(n) = x(n) + v(n) \tag{6.8}$$

where $v(n)$ has zero-mean with variance σ_v^2, and is independent of $u(n)$. The estimated ACF $r_y(\tau)$ of the noisy observation $y(n)$ can be computed as

$$r_y(\tau) = r_x(\tau) + r_v(\tau) + r_{vx}(\tau) + r_{xv}(\tau) \tag{6.9}$$

where the effect of cross-correlation terms are generally neglected considering that $x(n)$ and $v(n)$ are uncorrelated, therefore an estimate of $r_x(\tau)$ from $r_y(\tau)$ is obtained as

$$r_x(\tau) = \begin{cases} r_y(\tau) - \sigma_v^2 & \tau = 0 \\ r_y(\tau), & \tau \neq 0 \end{cases} \tag{6.10}$$

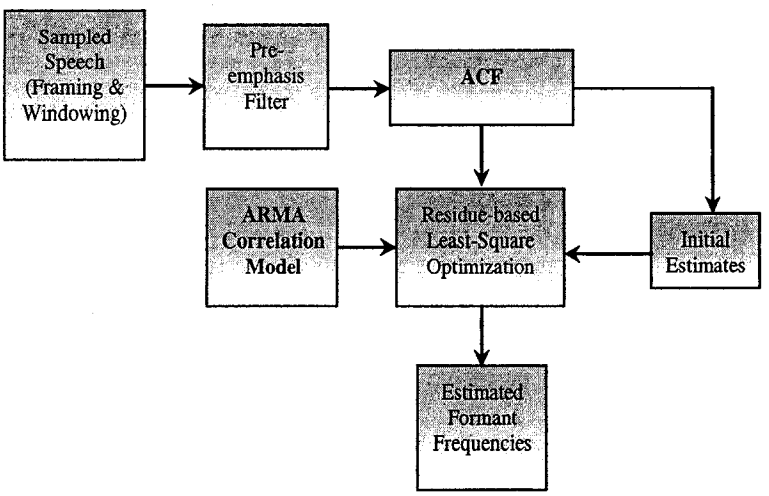


Figure 6.5: Block diagram of the proposed ARMA correlation model based formant estimation method.

However, under heavy noisy condition, estimation of $r_x(\tau)$ may cause significant error at all lags, resulting in poor pole estimates in the conventional correlation based methods. To alleviate this problem, the correlation-fitting based approach proposed in Chapter 2 is employed along with an adaptive RBLS algorithm for the estimation under a heavy noisy condition.

Fig. 6.5 shows a block-diagram explaining the the formant estimation method using the ARMAC model. In comparison to the block-diagram of the general framework involved in the proposed method as shown in Fig. 6.2, the main difference in Fig. 6.5 is the block that estimates the formant frequencies from the ACF $r_y(\tau)$ of the noisy observations. In $r_y(\tau)$, the effect of noise is severe mainly at $\tau = 0$. Hence, in order to reduce the effect of noise, $\tau > 0$ is considered.

The parameters of each component $F_l(\tau)$ of (6.7) with $\tau > 0$, are determined such that the total squared error between the $(l-1)$th residual function and $F_l(\tau)$ is

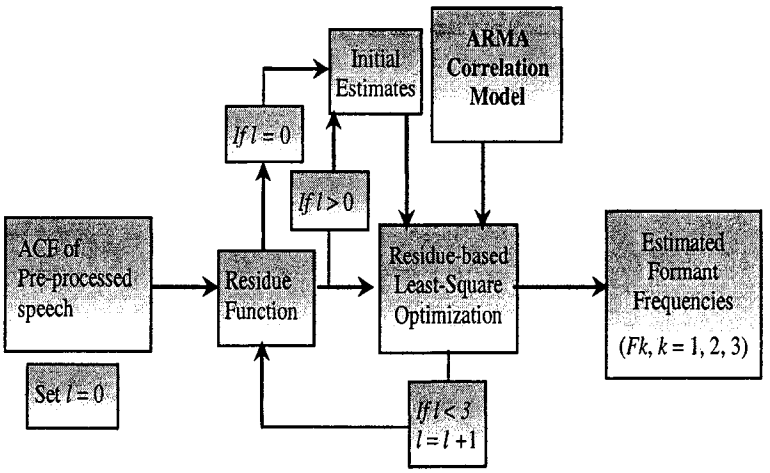


Figure 6.6: Adaptive RBLS algorithm involved in the proposed ARMA correlation model based formant estimation method.

minimized. We define the lth residual function as

$$\begin{aligned} \Re_l(\tau) &= \Re_{l-1}(\tau) - F_l(\tau);\ l = 1, 2, \ldots, \theta - 1; \\ \Re_0(\tau) &= r_y(\tau) \end{aligned} \tag{6.11}$$

Then, the objective function for the minimization problem becomes

$$J_l = \sum_{\tau=1}^{M-1} |\Re_{l-1}(\tau) - F_l(\tau)|^2,\ l = 1,\ 2,\ \ldots,\ \theta \tag{6.12}$$

The values for r_l and ω_l in the range $0 < r_l < 1$ corresponding to the global minimum of J_l are chosen as the estimate of the pole locations. Note that, in order to suppress the effects of the pitch from the first-formant estimation, an FFT pre-filtering is performed at the beginning of formant estimation. Fig. 6.6 shows a flow diagram explaining the adaptive RBLS algorithm estimating first three formants. In the adaptive RBLS method, K formant frequencies are sequentially determined from the K stages. The possible extreme ranges of the formants (ROF) are available in literature and are utilized to restrict the search space [7], [87]. The RBLS algorithm is

made adaptive with each stage by using the updated initial frequency estimates. At the first step ($l = 0$ in Fig. 6.6) of pre-filtering, frequency candidates for first formant (F1) are estimated from the smoothed spectral peaks. Candidates inside the desired region specified by the ROF are taken as the initial estimates and the frequency search is performed in their neighborhoods. For the initial estimates at the remaining steps ($l = 1, 2, ...$) of the adaptive RBLS algorithm, smoothed spectral peaks of the residue functions obtained for the corresponding steps are used. Pole magnitude can also be updated in each stage depending on the relative energy of residue function and the practical knowledge about the bandwidths of the different formants [7], [87].

6.4.2 Simulation Results

The proposed formant frequency estimation algorithm based on the ARMAC model is tested using various synthetic and natural vowels from the North-Texas vowel database [121] as well as some natural sentences from the TIMIT speech database [120]. Recently, a reliable reference database for the vocal-tract resonances of a large number of TIMIT sentences is reported in [147]. This vocal-tract resonances database is carefully used (keeping in mind the differences between VTR and formant frequencies [87]) to test the proposed method. For the performance comparison, we consider the LPC of order 14 [7] and the adaptive filter-bank (AFB) methods [83].

At first, results for three synthetic vowels /a/, /u/, and /i/ corrupted by additive white Gaussian noise are presented. Vowels are synthesized using the Klatt synthesizer with male and female pitch values, respectively, 120 Hz and 220 Hz, and 200 ms of duration. The formant estimation in the proposed method is conducted every 10 ms with a 20 ms window only for voiced frames. For the purpose of ARMAC model-fitting by using the RBLS algorithm, r_l is searched in the range $[0.8, 0.99]$

Table 6.1: Average RMSE (Hz) for synthetic vowels (Male)

Vowels		0 dB			5 dB		
		Prop.	LPC	AFB	Prop	LPC	AFB
/a/	F1	44.3	128.1	244.7	36.7	104.9	98.5
	F2	75.2	358.6	458.8	33.4	177.4	140.5
	F3	198.2	403.3	461.9	128.9	305.6	182.1
/u/	F1	109.3	294.2	292.1	50.6	96.2	143.4
	F2	205.4	711.2	720.1	57.4	232.2	204.1
	F3	252.1	774.7	620.2	176.0	376.8	317.0
/i/	F1	129.8	254.8	282.3	57.7	127.1	149.6
	F2	77.1	156.2	53.8	38.2	130.1	47.2
	F3	100.2	268.5	164.8	71.8	144.5	71.6

Table 6.2: Average RMSE (Hz) for synthetic vowels (Female)

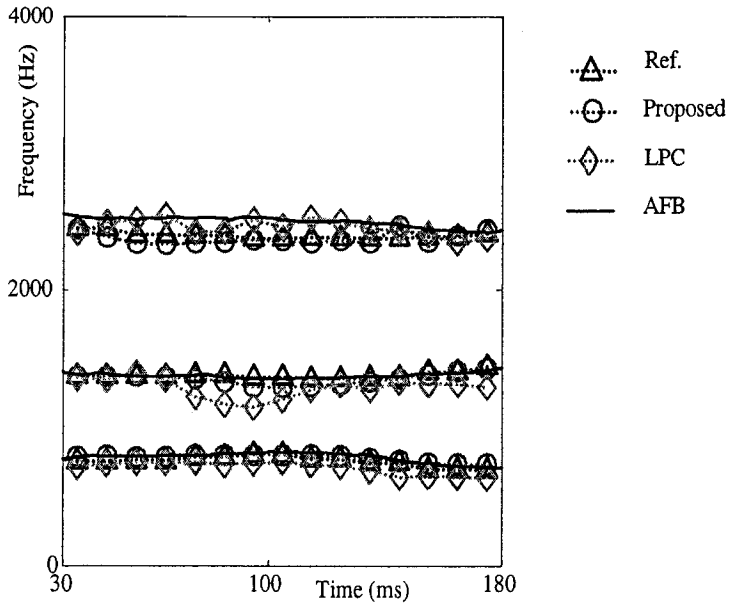
Vowels		0 dB			5 dB		
		Prop.	LPC	AFB	Prop	LPC	AFB
/a/	F1	98.6	185.2	265.5	100.0	122.9	278.3
	F2	77.4	186.0	205.0	60.3	124.3	116.2
	F3	80.3	92.16	80.7	73.8	81.3	72.9
/u/	F1	134.9	396.1	379.2	93.3	117.9	132.0
	F2	140.3	599.6	356.1	125.5	241.5	201.1
	F3	163.8	519.5	259.4	115.4	287.1	152.9
/i/	F1	153.2	329.8	490.2	77.7	203.0	233.8
	F2	179.3	537.4	319.2	107.1	119.2	142.6
	F3	51.7	244.3	47.5	34.2	75.7	25.4

for F3 and $[0.85, 0.99]$ for F2 and F1, and the search range for ω_l is chosen as 0.1π around the neighborhood of the initial estimates [93]. An acceptable level of estimation accuracy can be achieved with a search resolution of $\Delta r = 0.01$ for r_l and $\Delta\omega = 0.01\pi$ for ω_l with $0 < \omega_l < \pi$. The number of lags for the ACF is set to be $M = min(0.2N, T/2)$. Formant estimation error rate is calculated for voiced sounds only. We have computed the root-mean-square error (RMSE) at different noise levels, and in each level 20 independent trials are considered. Tables 6.1 and 6.2 show the average of RMSE values (Hz) obtained by different methods at SNR = 0 dB and 5 dB. Clearly, the performance of the proposed (Prop.) method is superior to that of the other two methods at both levels of SNR, and the RMSE for other methods increases significantly at SNR = 0 dB.

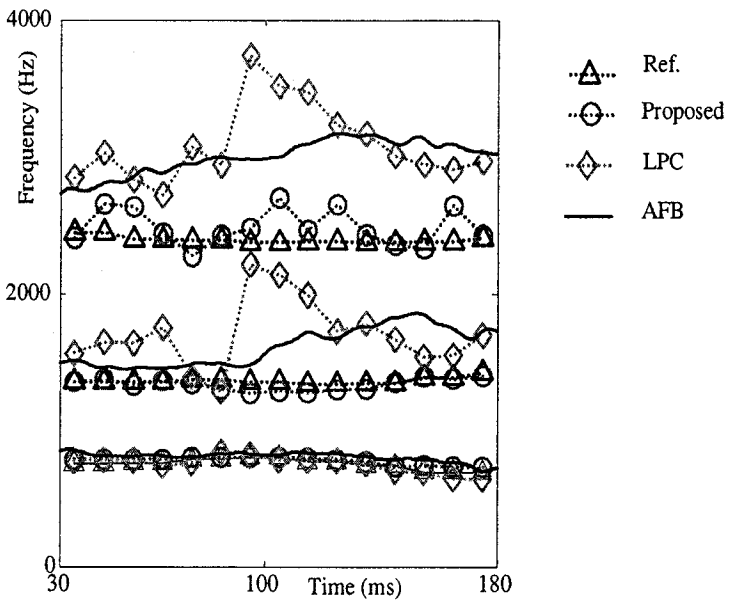
Next, four natural vowels /a/, /u/, /i/, and /e/ are taken from the North-Texas vowel database [121] with reference formant values. The vowels were contained in the words "hod", "hood", "heed", and "head". For the purpose of analysis, pitch periods are determined from the noise-free speech signals using autocorrelation method [7]. Estimated formant frequencies for a male vowel /a/ is considered as a white Gaussian noise are plotted in Fig. 6.7 at SNR = 20 and SNR = 0 dB. Similarly, estimated formant frequencies for a female vowel /i/ under white Gaussian noise are plotted in Fig. 6.8 at SNR = 20 and SNR = 0 dB. It is evident that the proposed method can track the formant frequencies quite accurately even at levels of SNR as low as 0 dB. Due to the high energy at the first formant (F1) level, F1 can be tracked well by all three methods even at 0 dB. However, due to the low energy at F3, presence of strong background noise makes F3-tracking very difficult. But the proposed method is still able to overcome this difficulty by employing the adaptive RBLS algorithm to extract the correlation model parameters giving the formant frequencies.

In Fig. 6.9 the effect of change in noise level (SNR) on average root-mean-square error (RMSE) is plotted for male natural vowel /a/ from SNR = 0 dB to 40 dB. The difference in RMSE values obtained by the proposed and other methods is quite high for F2 and F3. The result in the presence of multiple background competing speakers (Babble noise) is also presented in Fig. 6.10 to demonstrate the ability of the proposed method to deal with the environmental noise. The multiplicity of speakers produces a flatter short-term spectrum which has greater spectral and temporal modulation than white noise. Fig. 6.10 shows that the proposed method works well for both male and female speakers in background babble noise even at SNR = 0 dB.

Finally, we present the estimation results for a natural male utterance "Rob sat by the pond" which is taken from the TIMIT database (sampling frequency = 16 KHz).

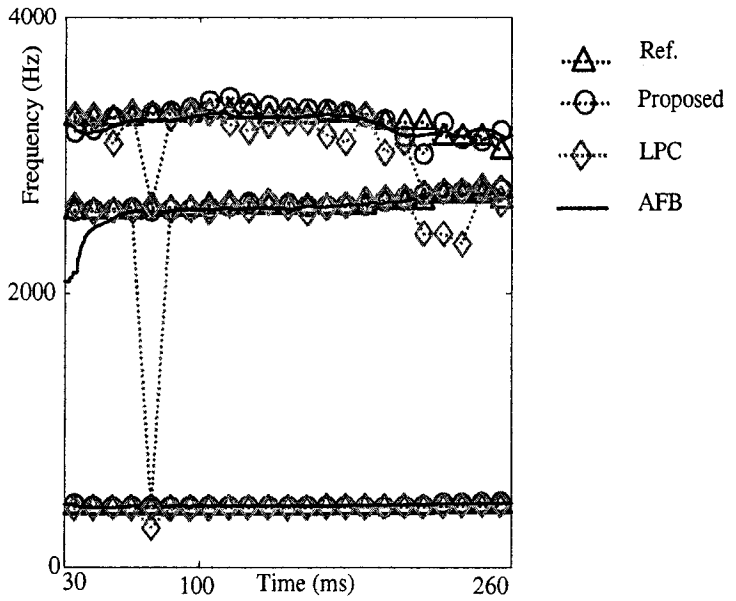


(a)

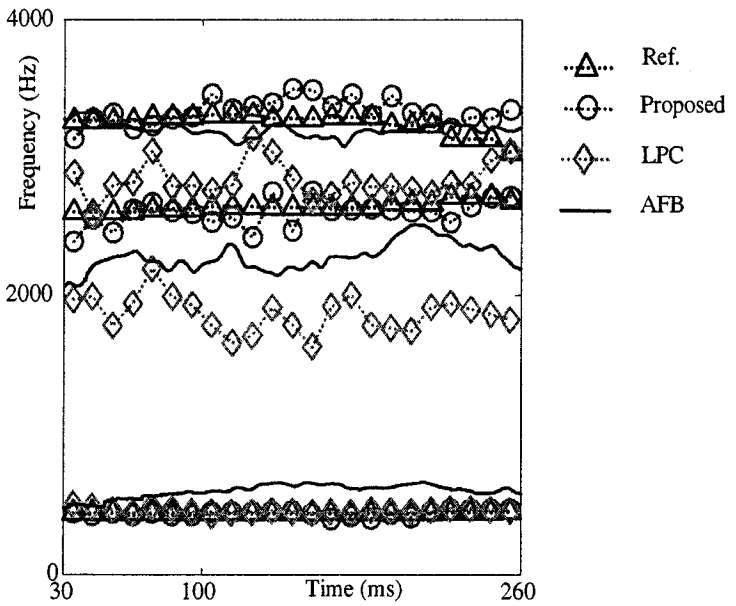


(b)

Figure 6.7: Formant tracks for male vowel /a/ in the presence of white noise at SNR levels of (a) 20 dB and (b) 0 dB.



(a)



(b)

Figure 6.8: Formant tracks for female vowel /i/ in the presence of white noise at SNR levels of (a) 20 dB and (b) 0 dB.

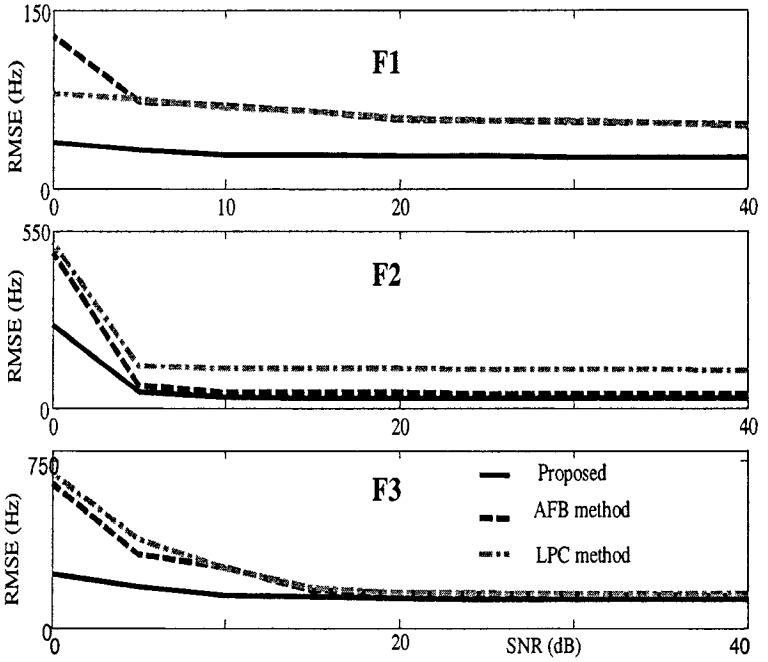
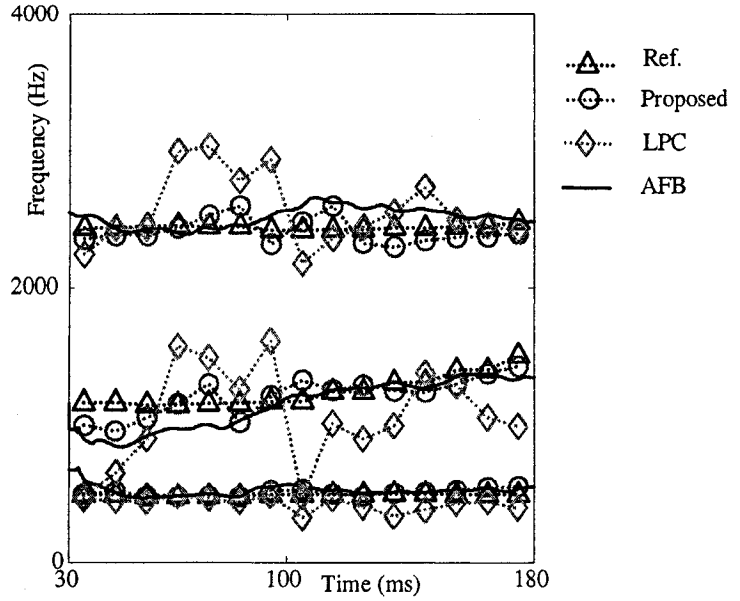
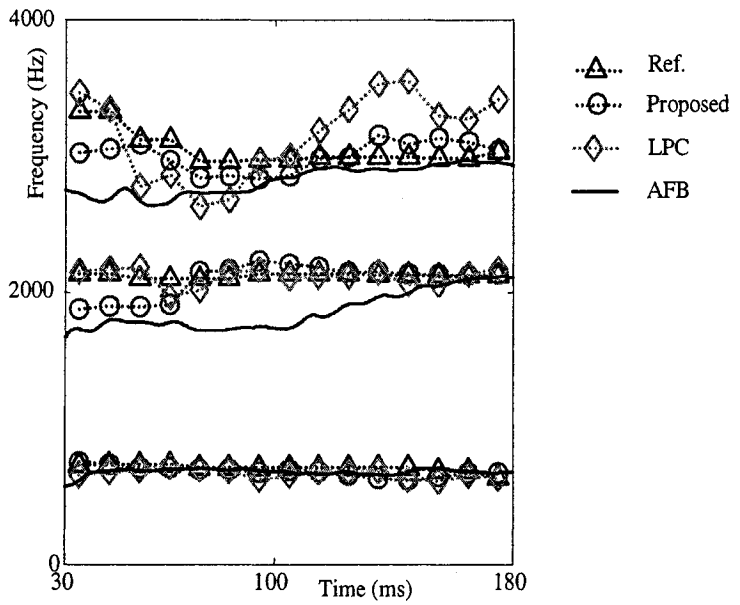


Figure 6.9: Effect of SNR on average RMSE (Hz) in the estimation of different formants (F1, F2, and F3).

First we perform the pre-emphasis operation on the FFT-filtered speech signal. Then every 10 ms, a 20 ms hamming window is applied to overlapping speech segment. In Fig. 6.11 (a), the reference formant frequencies are plotted on the spectrogram of noise-free speech [120]. The estimated formants by different methods at SNR = 5 dB in the presence of white Gaussian noise are plotted on the spectrogram of the noise-free signal to clearly show the formant trajectory during the voiced regions. For fair comparison voicing decisions are taken from the AFB method. Formant frequencies are estimated only in these voiced frames (dark line in the spectrogram) and in the spectrogram the interval between the two voiced frames are just end-point connected (dotted line) for Figs. 6.11(a) to 6.11(c). Since the AFB method works on sample by sample, it provides values also for those intervals. If we compare the estimation



(a)



(b)

Figure 6.10: Formant tracks at SNR = 0 dB in the presence of multi-talker babble noise for (a) male vowel /u/ and (b) female vowel /e/.

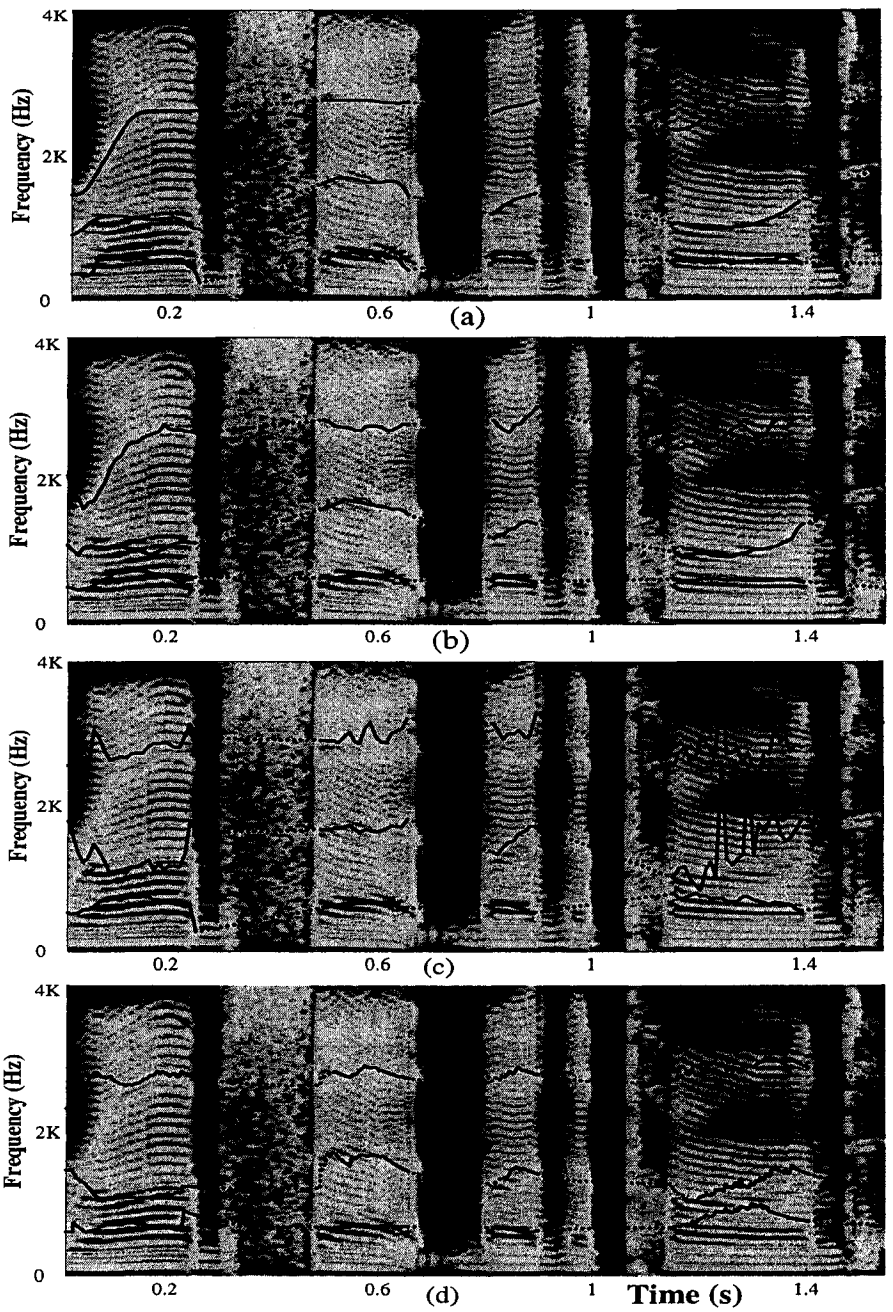


Figure 6.11: Formant estimation results for a male utterance "Rob sat by the pond" at SNR = 5 dB plotted on clean speech spectrogram. (a) Reference, (b) Proposed ARMAC model based method, (c) LPC method, and (d) AFB method.

accuracy in the voiced regions, it is evident that the proposed method can track the rise-fall pattern of different formants almost accurately except in the last segment accuracy in F3 tracking degrades, the reason is the same as explained earlier.

6.5 Ramp Cepstrum Model Based Formant Estimation

6.5.1 Proposed Method

In Chapter 3, a ramp-cepstrum model of OSACF of the output of an AR system excited by a periodic impulse-train has been derived. For a voiced speech segment within a short duration of time, the proposed AR ramp-cepstrum (ARRC) model can be used to extract the poles of the overall vocal tract system. Since formant frequencies can be estimated from the poles of the overall vocal-tract system, an accurate estimate of the ARRC model parameters from noise-corrupted speech signal is essential to obtain formant frequencies. In this section a ramp-cepstrum model-fitting based approach is employed in conjunction with the adaptive RBLS optimization algorithm in order to obtain an accurate formant frequency estimation from noise-corrupted output observations.

The OSACF of noise-free speech can be obtained as

$$r_x(\tau) = \begin{cases} \rho_x(\tau), & \tau > 0 \\ 0.5\rho_x(\tau), & \tau = 0 \\ 0, & \tau < 0 \end{cases} \tag{6.13}$$

where $\rho_x(\tau)$ is the two-sided ACF of $x(n)$. From the Fourier transform $R_x(e^{j\omega})$ of $r_x(\tau)$, in Chapter 3, the complex cepstrum for $n > 0$ has been obtained as

$$\mu_x(n) = F^{-1}\{\ln(\Re e[R_x(e^{j\omega})])\} = \sum_{i=1}^{P} \frac{p_i^n}{n}, n > 0 \tag{6.14}$$

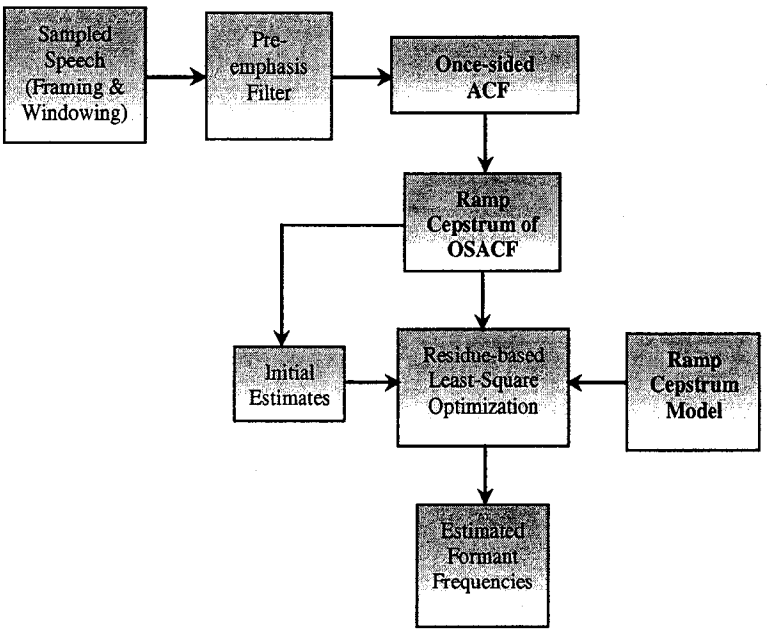


Figure 6.12: Block diagram of the proposed ramp-cepstrum model based formant estimation method.

The ramp-cepstrum (RC) model that was proposed in Chapter 3 is given by

$$\psi_x(n) = n\mu_x(n) = \sum_{i=1}^{\gamma} \beta(\omega_i) r_i^n \cos(\omega_i n), 0 < n < T \tag{6.15}$$

Note that γ in (6.15) is the number of pairs of complex conjugate poles with $\beta(\omega_i) = 2$. In the presence of noise, the RC of $r_y(\tau)$ can be expressed as

$$\psi_y(n) = \psi_x(n) + \psi_w(n), 0 < n < T \tag{6.16}$$

Here, $\psi_w(n)$ is the error introduced due to the noise. Each of the γ component functions in (6.15) is estimated sequentially from M_c instances of $\psi_y(n)$, with $M_c < T$ by using the adaptive RBLS algorithm as described in 6.4.2.

In Fig. 6.12, a block-diagram explaining the the formant estimation method using the ARRC model is presented. In comparison to the block-diagram of the general framework involved in the proposed method as shown in Fig. 6.2, the main difference

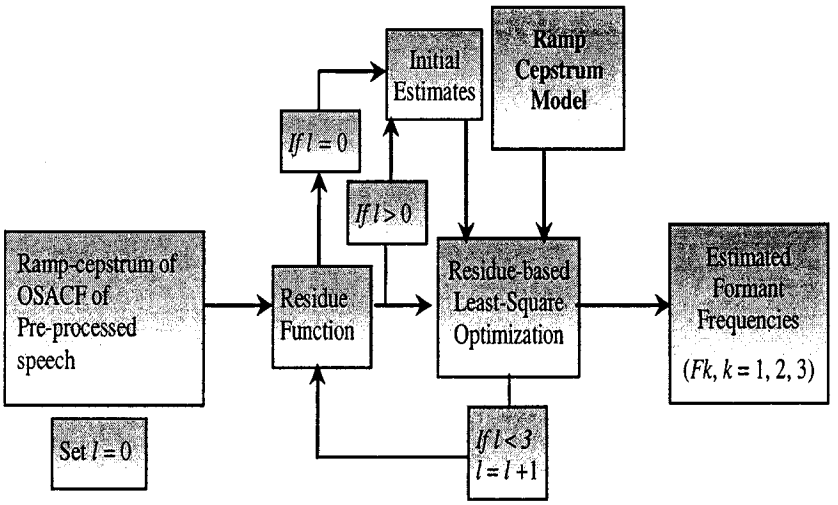


Figure 6.13: Adaptive RBLS algorithm involved in the proposed ramp-cepstrum model based formant estimation method.

in Fig. 6.12 is the block that estimates the formant frequencies from the ACF $r_y(\tau)$ of the noisy observations. In $r_y(\tau)$, the effect of noise is severe mainly at $\tau = 0$. Hence, in order to reduce the effect of noise, we exclude $r_y(0)$ to compute $\psi_y(n)$ at a low SNR, which significantly reduces the effect of noise. The objective function is defined as

$$J_l = \sum_{n=1}^{M_c} |\Re_{l-1}(n) - \beta(\omega_l) r_l^n \cos(\omega_l n)|^2, l = 1, \ldots, \gamma \tag{6.17}$$

where the residual function $\Re_l(n)$ is updated as follows

$$\begin{aligned} \Re_0(n) &= \psi_y(n) \\ \Re_l(n) &= \Re_{l-1}(n) - \beta(\omega_l) r_l^n \cos(\omega_l n),\ l = 1, \ldots, \gamma - 1 \end{aligned} \tag{6.18}$$

Fig. 6.13 presents a block diagram explaining the adaptive RBLS algorithm in order to estimate the first three formants based on the ARRC model. The values for r_l and ω_l in the range $0 < r_l < 1$ corresponding to the global minimum of J_l are chosen as the estimate of the pole locations. In the adaptive RBLS method, formant

frequencies are sequentially determined, giving one formant in each step. At the first step ($l = 0$ in Fig. 6.13), frequency candidates for F1 are estimated from the smoothed sectral peaks of the noisy OSACF. Candidates inside the desired region specified by the ROF are taken as the initial estimates and the frequency search is performed in their neighborhoods. For the initial estimates at the remaining steps ($l = 1, 2, ...$) of the adaptive RBLS algorithm, smoothed spectral peaks of the residue functions of corresponding steps are used.

6.5.2 Simulation Results

The proposed method of formant frequency estimation based on the ARRC model is tested using various synthetic and natural vowels from the North-Texas vowel database [121], and some natural sentences from the TIMIT speech database [120]. [147]. For the performance comparison we consider the LPC of order 14 [7] and adaptive filter-bank (AFB) methods [83]. In the proposed method, the formant estimation is conducted every 10 ms with a 20 ms window only for voiced frames. Moreover, formant estimation error is computed for voiced sounds only. We have computed the root-mean-square error (RMSE) at different noise levels, and in each level 20 independent trials are considered. For the purpose of ramp-cepstral model-fitting by using the RBLS algorithm, r_l is searched in the range $[0.8, 0.99]$ for F3 and $[0.85, 0.99]$ for F2 and F1, and the search range for ω_l is chosen as 0.1π around the neighborhood of the initial estimates [93]. Search resolutions are set to $\Delta r = 0.01$ and $\Delta\omega = 0.01\pi$. The number of ramp-cepstral instances is set to be $N_c = min(0.2N, T/2)$.

In Tables 6.3 and 6.4, the estimation performance for three synthetic vowels $/a/$, $/u/$, and $/i/$ in the presence of white Gaussian noise is presented. Tables 6.3 and 6.4 shows the average of RMSE values (Hz) obtained by different methods at SNR

Table 6.3: Average RMSE (Hz) for synthetic vowels (Male)

Vowels		0 dB			5 dB		
		Prop.	LPC	AFB	Prop	LPC	AFB
/a/	F1	64.1	128.1	244.7	55.2	104.9	98.5
	F2	102.2	358.6	458.8	44.1	177.4	140.5
	F3	209.2	403.3	461.9	185.9	305.6	182.1
/u/	F1	112.3	294.2	292.1	55.3	96.2	143.4
	F2	263.4	711.2	720.1	106.5	232.2	204.1
	F3	252.1	774.7	620.2	179.5	376.8	317.0
/i/	F1	131.7	254.8	282.3	77.4	127.1	149.6
	F2	109.1	156.2	53.8	39.7	130.1	47.2
	F3	119.2	268.5	164.8	73.2	144.5	71.6

Table 6.4: Average RMSE (Hz) for synthetic vowels (Female)

Vowels		0 dB			5 dB		
		Prop.	LPC	AFB	Prop	LPC	AFB
/a/	F1	99.7	185.2	265.5	83.9	122.9	278.3
	F2	98.2	186.0	205.0	61.6	124.3	116.2
	F3	80.3	92.16	80.7	72.2	81.3	72.9
/u/	F1	58.5	396.1	379.2	97.4	117.9	132.0
	F2	160.3	599.6	356.1	135.8	241.5	201.1
	F3	189.8	519.5	259.4	145.4	287.1	152.9
/i/	F1	196.2	329.8	490.2	133.7	203.0	233.8
	F2	201.3	537.4	319.2	109.1	119.2	142.6
	F3	53.1	244.3	47.5	37.4	75.7	25.4

= 0 dB and 5 dB. It is observed that the RMSE values obtained by other methods increase significantly at SNR = 0 dB. The formant frequency estimation performance of the proposed (Prop.) method is significantly better than that of the other two methods for both levels of SNR.

The estimation performance of the proposed method is tested for different natural vowels uttered by several speakers taken from the North-Texas vowel database [121] with reference formant values. In Table 6.5, the estimation performance of different methods for a male vowel /a/ and a female vowel /i/ in the presence of white Gaussian noise is presented in terms of mean and standard deviation (shown in the parenthesis) at SNR = 5 dB. It can be seen that the formant frequency estimated by the proposed method, as compared to that by other methods, is more accurate and consistent for

Table 6.5: Estimated mean and standard deviation for natural vowels

Fi	Male (/a/)				Female (/i/)			
	Ref.	Prop.	LPC	AFB	Ref.	Prop	LPC	AFB
F1	754	789	815	826	435	427	422	451
	(33)	(41)	(58)	(51)	(3)	(14)	(17)	(28)
F2	1369	13.89	14.16	1401	2638	2584	1990	2284
	(22)	(67)	(85)	(72)	(35)	(81)	(126)	(141)
F3	2402	2447	2551	2516	3250	3317	2831	2976
	(29)	(96)	(123)	(101)	(70)	(116)	(128)	(116)

both male and female speakers. As mentioned before, the low level of energy in the F3 band and the presence of strong background noise make F3-estimation difficult. However, by employing the adaptive RBLS algorithm in ramp-cepstrum model-fitting, the proposed method overcome such problems.

In Fig. 6.14, the effect of noise on the estimation errors in terms of average RMSE is plotted for a male natural vowel /a/ from SNR = 0 dB to 40 dB. The difference in the RMSE values obtained by the proposed and other methods is quite high for F3.

The estimation accuracy of different methods in the presence Babble noise is shown in Fig. 6.15. It is observed that the proposed method works well for both male and female speakers in background babble noise even at SNR = 0 dB.

In Fig. 6.16, the formant estimation results are presented in a similar fashion as it has been shown in Section 6.4.2. In this figure, a natural male utterance "Rob sat by the pond" which is extracted from the TIMIT database is considered in the presence of white noise at SNR = 5 dB. In Fig. 6.16 (a) the reference formant frequencies are plotted on the spectrogram of noise-free speech. As explained in Section 6.4.2, formant frequencies are estimated only in the voiced frames. From this figure, it is clearly observed that the proposed method yields a better estimation accuracy even in some difficult regions where the formant energy is very low (e.g., F3) or a rise-fall

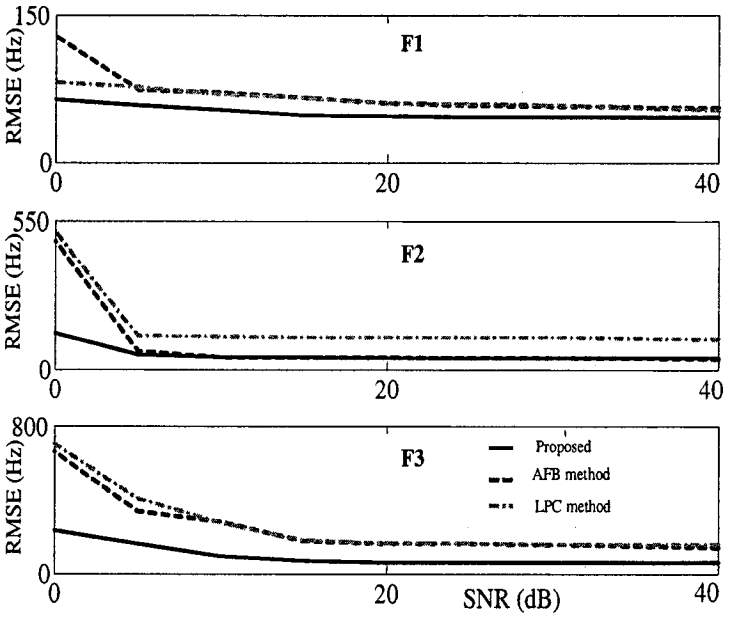
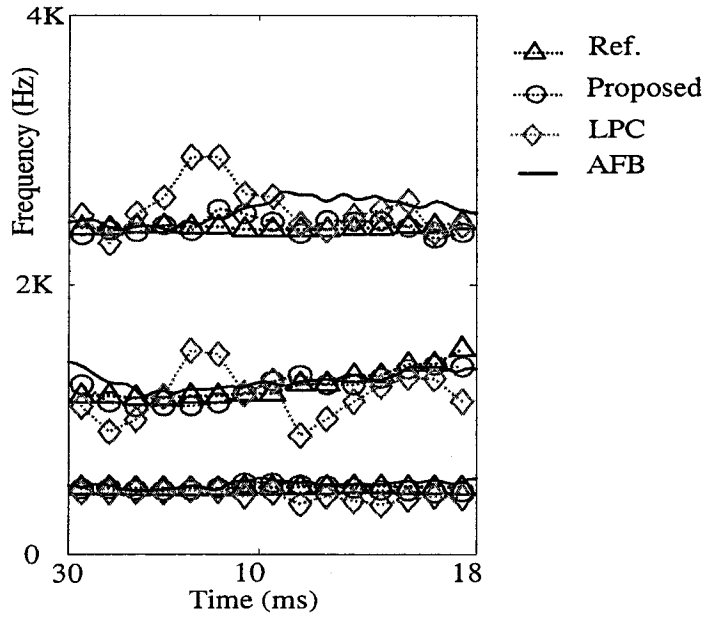


Figure 6.14: Effect of SNR on average RMSE (Hz) in the estimation of different formants (F1, F2, and F3).

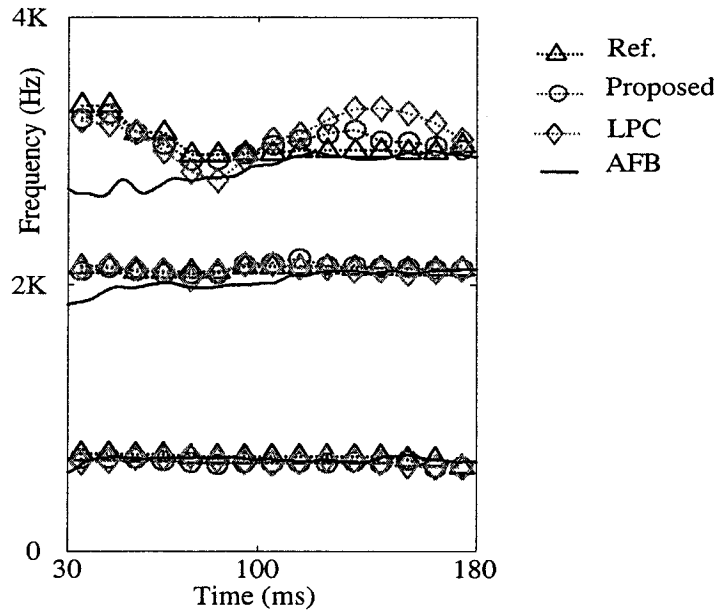
pattern appears.

6.6 Ramp Cosine Cepstrum Model Based Formant Estimation

In Chapter 4, a ramp cosine cepstrum (RCC) model of OSACF of the output of an AR system excited by a periodic impulse-train has been derived. Since, formant frequencies can be estimated from the poles of the overall vocal-tract system, an accurate estimate of the RCC model parameters from noise-corrupted speech signal is essential to obtain formant frequencies. Recently, it has been reported that instead of conventional ACF, if repeated ACF is employed under a heavy noisy condition, a better identification performance is obtained [126], [128], [137], [148]–[150]. In this



(a)



(b)

Figure 6.15: Formant tracks at SNR = 0 dB in the presence of babble noise for (a) male vowel /u/ and (b) female vowel /e/.

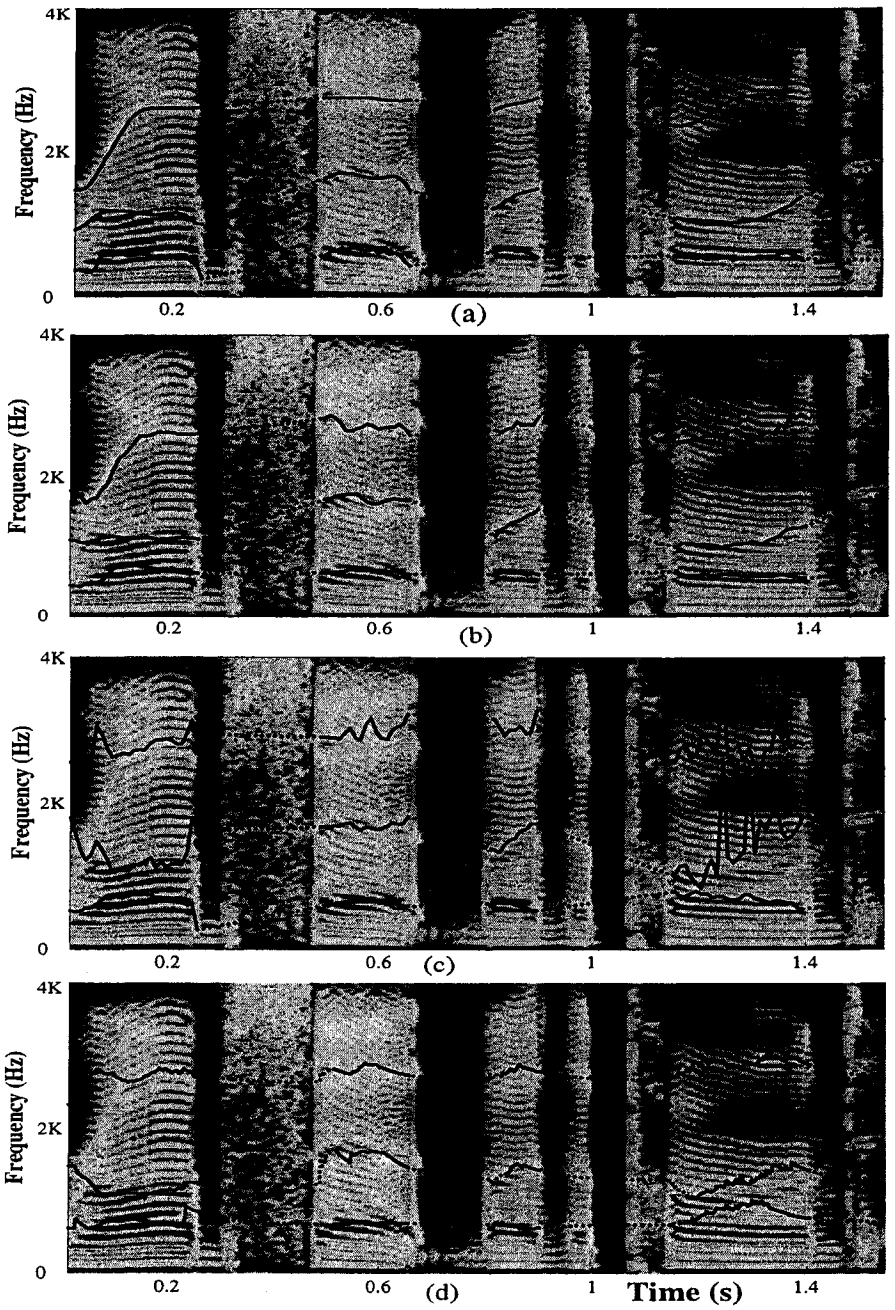


Figure 6.16: Formant estimation results for a male utterance "Rob sat by the pond" at SNR = 5 dB plotted on clean speech spectrogram; (a)Reference, (b) Proposed ramp cepstrum model based method, (c) LPC method, and (d) AFB method.

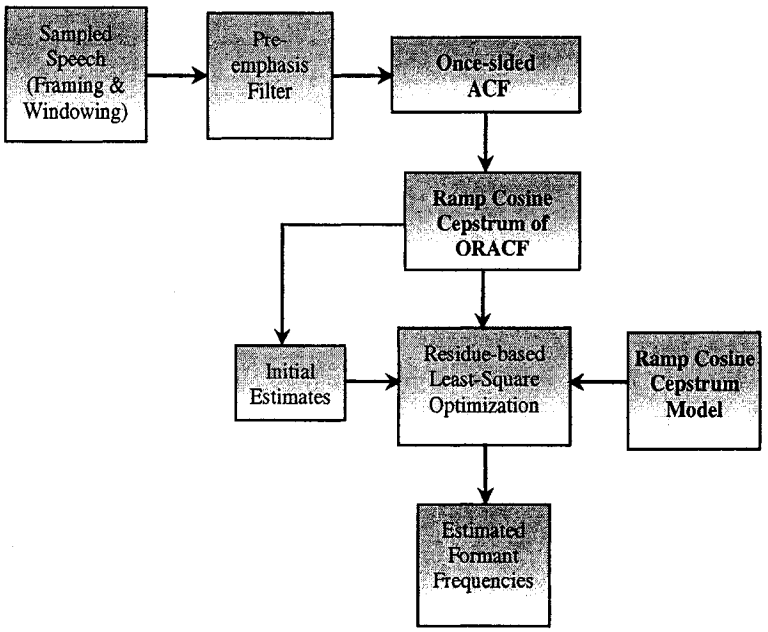


Figure 6.17: Block diagram of the proposed ramp cosine cepstrum model based formant estimation method.

section, we intend to employ a once-repeated ACF in the development of an RCC model. The effect of repeating the ACF is also addressed in this section. A new model is developed for the the ramp cosine cepstrum of the once repeated ACF (ORACF). In comparison to the OSACF, the ORACF offers more noise robustness. Based on the new model, a ramp-cepstrum model-fitting using the adaptive RBLS optimization algorithm is performed in order to obtain an accurate formant frequency estimation from noise-corrupted output observations.

6.6.1 Proposed Method

A block diagram showing the main steps of the proposed ramp cosine cepstrum model based formant estimation scheme is shown in Fig. 6.17. In comparison to the block-diagram of the general framework involved in the proposed method as shown in Fig.

6.2, the main difference in Fig. 6.17 is the block that estimates the formant frequencies from the ACF $r_y(\tau)$ of the noisy observations. Here, an RCC model is developed in a manner similar to that obtained in Chapter 4. In order to obtain a better noise immunity, instead of conventional ACF, a once-repeated ACF (ORACF) of noise-free speech $x(n)$, defined as

$$\psi_x(m) = \begin{cases} \phi_x(m), & m > 0 \\ 0.5\phi_x(m), & m = 0 \\ 0, & m < 0 \end{cases} \tag{6.19}$$

is employed where $\phi_x(m)$ is obtained by repeating the autocorrelation operation on a conventional ACF $r_x(m)$ defined in (6.4). Now $\phi_x(m)$ can be expressed as

$$\phi_x(m) = \psi_x(m) + \psi_x(-m). \tag{6.20}$$

It can be shown that $\psi_x(m)$ retains the pole-preserving property of $r_x(m)$. Moreover, under a noisy condition, the ORACF of the observed data offers an advantage of higher noise immunity than the conventional ACF. Using the definition given in (4.15), the cosine transform of a real signal $\{\psi_x(m)\}_{m=0}^{N-1}$ is obtained as

$$c_{\psi_x}(m) = \mathcal{F}_c^{-1}[\ln[\mathcal{F}_c[\psi_x(m)]]] \tag{6.21}$$

where $\mathcal{F}_c^{-1}[\cdot]$ denotes the inverse operator of the cosine transform defined in (4.16). Since $r_x(m)$ is the ACF of $x(n)$, the Fourier transform of $r_x(m)$, say $R_x(e^{j\omega})$, equals $|X(e^{j\omega})|^2$ and for the linear time-invariant system with the transfer function $H(z)$, $R_x(e^{j\omega})$ can be written as

$$R_x(e^{j\omega}) = \left|H(e^{j\omega})\right|^2 R_w(e^{j\omega}) \tag{6.22}$$

where $R_w(e^{j\omega})$ is the Fourier transform of $r_w(\tau)$. Similarly, the ORACF $\phi(m)$ has a Fourier transform $\Phi(e^{j\omega}) = |R(e^{j\omega})|^2$. Hence, $c_{\psi_x}(m)$ in (6.21) can be expressed as

$$c_{\psi_x}(m) = 2\mathcal{F}_c^{-1}[\ln[H(e^{j\omega})]] + 2\mathcal{F}_c^{-1}[\ln[H(e^{-j\omega})]]$$
$$+\mathcal{F}_c^{-1}[\ln[|R_w(e^{j\omega})|^2]] + \mathcal{F}_c^{-1}\left[\ln\left[\frac{1}{2}\right]\right] \tag{6.23}$$

It can be observed from (6.23) that the effect of the input excitation $w(n)$ can be made additive by using the homomorphic deconvolution, no matter which type of input excites the system. From (6.1), $\ln[H(z)]$ can be expanded as

$$\ln[H(z)] = -\sum_{i=1}^{M} \ln\left(1 - p_i z^{-1}\right) = \sum_{i=1}^{M}\sum_{n=1}^{\infty} \frac{p_i^n}{n} z^{-n} \tag{6.24}$$

where $|z| > |p_i|$. For a real and minimum phase $h(n)$, the inverse cosine transform of $\ln[\mathcal{F}[h(n)]]$ can be obtained as

$$\begin{aligned} \mathcal{F}_c^{-1}[ln[H(e^{j\omega})]] &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{i=1}^{M}\sum_{m=1}^{\infty} \frac{p_i^m}{m} e^{-j\omega m} cos\omega m \; d\omega \\ &= \frac{1}{2}\sum_{i=1}^{M}\sum_{m=1}^{\infty} \frac{p_i^m}{m}, \; m > 0 \end{aligned} \tag{6.25}$$

Similarly, for $\ln[H(e^{-j\omega})]$ we obtain

$$\mathcal{F}_c^{-1}[ln[H(e^{-j\omega})]] = \frac{1}{2}\sum_{i=1}^{M}\sum_{m=1}^{\infty} \frac{p_i^m}{m}, \; m > 0 \tag{6.26}$$

For voiced segments, the excitation is the periodic impulse-train $\{w_i(n)\}_{n=0}^{N-1}$ with a period T. It can be shown that the third term on the right side of (6.23) for $w_i(n)$ exhibits non-zero values at the origin and at multiples of T for $m \geq 0$. It then reduces to

$$\mathcal{F}_c^{-1}[\ln[|R_w(e^{j\omega})|^2]] = 0, \quad 0 < m < T \tag{6.27}$$

Hence, $c_{\psi_x}(m)$ in (4.19) can be expressed as

$$c_{\psi_x}(m) = 2\sum_{i=1}^{M} \frac{p_i^m}{m}, \quad 0 < m < T \tag{6.28}$$

In order to overcome the rapid decay in $c_{\psi_x}(m)$, we propose an easy-to-handle ramp cosine cepstrum (RCC) for the ORACF of $x(n)$ which is defined as

$$\chi_x(m) = mc_{\psi_x}(m) = 2\sum_{i=1}^{M} p_i^m, \quad 0 < m < T \tag{6.29}$$

Considering the complex conjugate poles, (6.29) can be expressed as

$$\chi_x(m) \;=\; \sum_{i=1}^{\kappa} \alpha(\omega_i) r_i^m \cos(\omega_i m), 0 < m < T \tag{6.30}$$

where κ is the number of complex conjugate pole pairs and $\alpha(\omega_i) = 4$. The ramp cosine cepstrum (RCC) model of the ORACF of $x(n)$ can be implemented using DCT-IDCT as

$$c_{\psi_x}(m) = IDCT[\ln(DCT[\psi_x(m)])], \; m > 0 \tag{6.31}$$

The detailed implementation of RCC for OSACF as described in Section 4.2.3 can be readily applicable for the case of ORACF.

In the presence of additive noise $v(n)$, the ACF of the noisy observation $y(n)$ can be expressed as

$$r_y(m) \;=\; r_x(m) + r_n(m) = r_x(m) + \{r_v(m) + r_{xv}(m) + r_{vx}(m)\} \tag{6.32}$$

For a finite data length, the effect of $r_n(m)$ cannot be neglected, especially when the SNR is very low. In order to reduce the effect of noise, we propose to use the ORACF $\psi_y(m)$ computed from $r_y(m)$. In this case, the strength of the ACF of $r_n(m)$ as well as the cross-correlation terms (between $r_n(m)$ and $r_x(m)$) is significantly reduced in comparison to the ACF of $r_x(m)$. In Fig. 6.18, normalized ACFs of a noise-free and noisy AR signal are plotted together with the ORACF of the noisy signal. In the presence of noise, the ACF of the noisy signal is severely degraded at all lags.

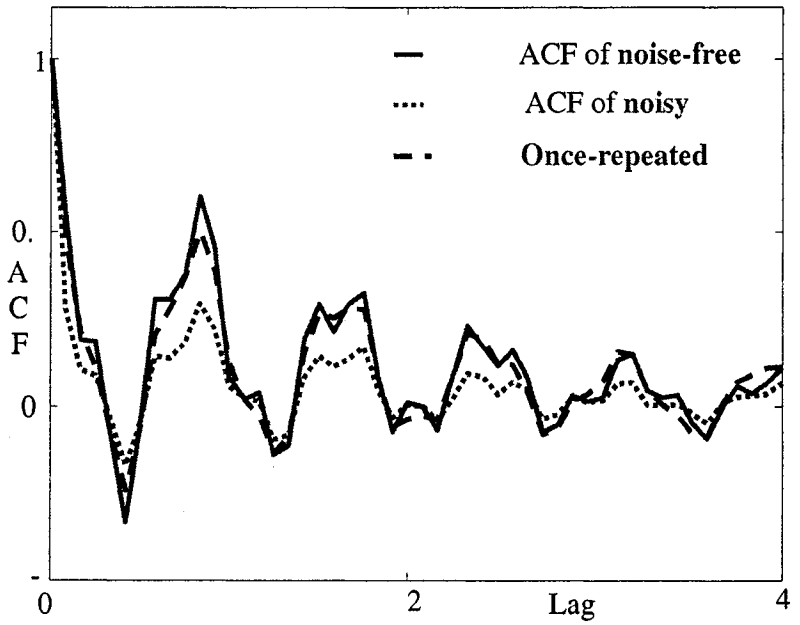


Figure 6.18: Effect of once-repeated ACF in noise.

However, it is clearly observed that the ORACF closely matches with the noise-free ACF. Hence, some system identification methods have been proposed which employ the repeated ACF for the identification of the AR part [126], [128], [137], [148]–[150]. In order to further reduce the effect of noise, the zero lag of $r_x(m)$ is made to zero before computing the ORACF. Finally, we switch to the cepstral domain where the logarithmic smoothing helps in retaining the shape of the RCC model under heavy noisy conditions. The RCC of the ORACF of noisy speech $y(n)$ can be expressed as

$$\chi_y(m) = \chi_x(m) + \chi_\epsilon(m), \quad m > 0 \tag{6.33}$$

Here, $\chi_\epsilon(m)$ is the error due to the noise that must be significantly reduced.

Fig. 6.19 presents a block diagram explaining the adaptive RBLS algorithm in order to estimate the first three formants based on the new RCC model. The values for r_l and ω_l in the range $0 < r_l < 1$ corresponding to the global minimum of J_l are

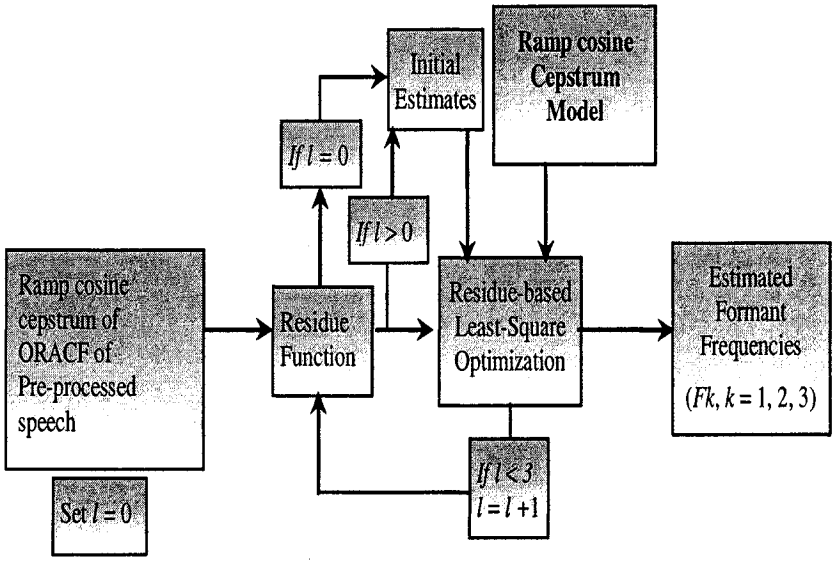


Figure 6.19: Adaptive RBLS algorithm involved in the proposed ramp cosine cepstrum model based formant estimation method.

chosen in a similar fashion as explained before. Each of the κ component functions is estimated sequentially from N_c ($< T$) instances of $\chi_y(m)$. The objective function is defined as

$$J_l = \sum_{n=1}^{N_c} |\Re_{l-1}(n) - \alpha(\omega_l) r_l^n \cos(\omega_l n)|^2 \,, \; l = 1, 2, \ldots, \kappa \tag{6.34}$$

where the residual function $\Re_l(n)$ is updated as follows

$$\begin{aligned} \Re_0(n) &= \chi_y(n) \\ \Re_l(n) &= \Re_{l-1}(n) - \alpha(\omega_l) r_l^n \cos(\omega_l n), \; l = 1, \ldots, \kappa - 1 \end{aligned} \tag{6.35}$$

The values of $\{r_l\}$ and $\{\omega_l\}$, corresponding to the global minimum of J_l, are selected to compute the estimate of lth formant. Thus, different formant frequencies are sequentially determined employing the adaptive RBLS as described in the ramp cepstrum model based formant estimation method.

6.6.2 Simulation Results

In this section, a number of simulations is carried out in order to demonstrate the effectiveness of the proposed method of estimating formant frequencies in the presence of noise. The proposed algorithm has been tested using different synthetic vowels synthesized using the Klatt synthesizer [7], natural vowels from the North-Texas vowel database [121], and some natural sentences from the TIMIT speech database with their reported formant references [147]. For the purpose of performance comparison in terms of formant estimation accuracy, the 14th order LPC and the adaptive filter-bank (AFB) methods have been considered [83].

In the adaptive RBLS algorithm, r_l is searched in the range $[0.8, 0.99]$. The search range for ω_l is chosen as 0.1π around the neighborhood of the initial estimates [93]. The initial estimates of ω_l are obtained from the location of the peaks of the smoothed DCT of the ORACF of $y(n)$ with $r_y(0) = 0$. The search range for ω_l can be narrowed down based on the knowledge of the pole-zero locations of a particular phoneme [7], [93]. Search resolutions of $\Delta r = 0.01$ and $\Delta\omega = 0.01\pi$ are used for r_l and ω_l, respectively. The number of ramp-cepstrum samples is taken as $N_c = \min(4P, T/2)$ where the pitch (T) can be estimated using the autocorrelation method [7].

First, the results for three synthetic vowels /a/, /u/, and /i/ under white Gaussian noise are presented. Vowels with duration of 200 ms are synthesized using the Klatt synthesizer considering the pitch values of 120 Hz and 220 Hz, respectively, for male and female speakers. Formant estimation is performed in every 10 ms with a 20 ms window only for voiced frames. In Tables 6.6 and 6.7, the estimated %RMSE (Hz) is shown for the three synthesized vowels at SNR = 0 dB and 5 dB for male and female sounds, respectively. It is clearly observed that the proposed method is able

Table 6.6: %RMSE (Hz) for synthetic vowels (Male)

Vowels		0 dB			5 dB		
		Prop.	LPC	AFB	Prop	LPC	AFB
/a/	F1	7.17	16.98	36.05	5.37	12.93	12.23
	F2	6.53	23.82	41.28	2.23	12.36	10.87
	F3	5.81	14.43	18.61	3.49	10.54	8.33
/u/	F1	17.29	42.36	51.49	11.81	23.99	27.64
	F2	10.51	36.61	34.46	7.35	15.25	12.81
	F3	10.62	23.59	21.27	6.46	13.92	11.31
/i/	F1	15.74	24.83	31.58	10.61	17.01	16.25
	F2	3.84	06.46	4.18	2.06	4.43	3.33
	F3	4.59	10.17	7.75	3.15	6.39	3.97

to reduce %RMSE significantly at both levels of SNR. The RMSE for other methods increases significantly at SNR = 0 dB.

Some natural vowels /a/, /u/, /i/, and /e/, contained in words "hod", "hood", "heed", and "head" respectively, are taken from the North-Texas database [121]. Estimated formant frequencies at different time instances are plotted in Figs. 6.20(a) and 6.20(b), respectively, for a male sound /a/ and a female sound /e/ at SNR = 0 dB. It is observed that the proposed method is capable of estimating all three formants quite accurately for both male and female speakers. In Fig. 6.21, the effect of the change in noise level (SNR) on the average RMSE is plotted for a male natural vowel /a/ from SNR = 0 to 40 dB at steps of 5 dB. The difference in RMSE values between the proposed and other methods is quite high for $F3$.

The performance in the presence of Babble noise is presented in Fig. 6.22 to demonstrate the ability of the proposed method to deal with the environmental noise. It is observed that the proposed method works well for both male and female speakers in background babble noise even at SNR = 0 dB.

In Fig. 6.23, estimation results for a natural utterance "*Rob sat by the pond*" uttered by an adult male are given. In Fig. 6.23 (a) the reference formant frequencies are plotted on the spectrogram of noise-free speech [147]. The formants estimated

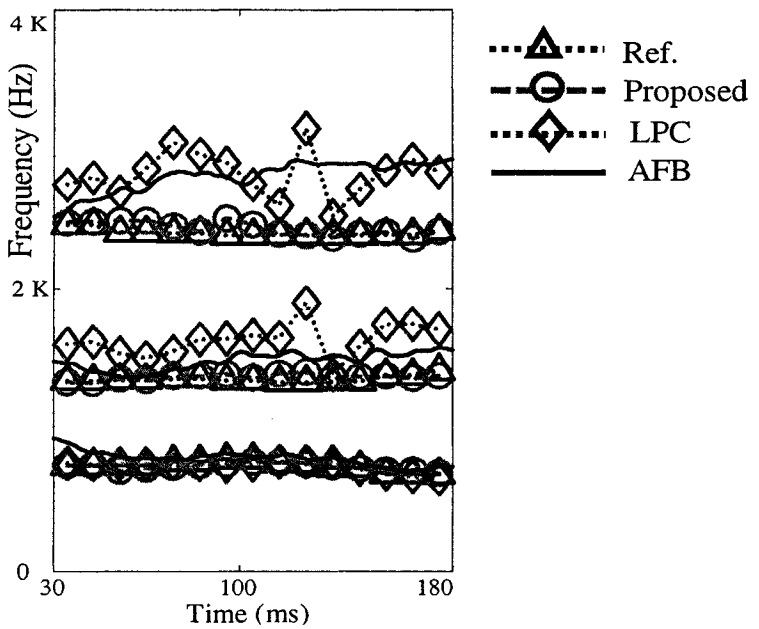
Table 6.7: %RMSE (Hz) for synthetic vowels (Female)

Vowels		0 dB			5 dB		
		Prop.	LPC	AFB	Prop	LPC	AFB
/a/	F1	7.84	11.97	17.19	3.51	5.27	15.67
	F2	5.15	15.69	15.21	3.17	8.76	7.49
	F3	2.78	4.58	3.97	2.02	2.35	2.28
/u/	F1	10.02	21.51	27.26	9.46	15.97	18.28
	F2	8.28	24.04	18.53	6.21	12.05	10.37
	F3	5.94	12.97	9.51	4.12	7.57	5.18
/i/	F1	16.59	37.81	35.19	10.18	18.69	20.08
	F2	7.43	19.65	14.05	3.52	9.51	6.28
	F3	2.54	7.93	3.54	1.97	2.46	2.41

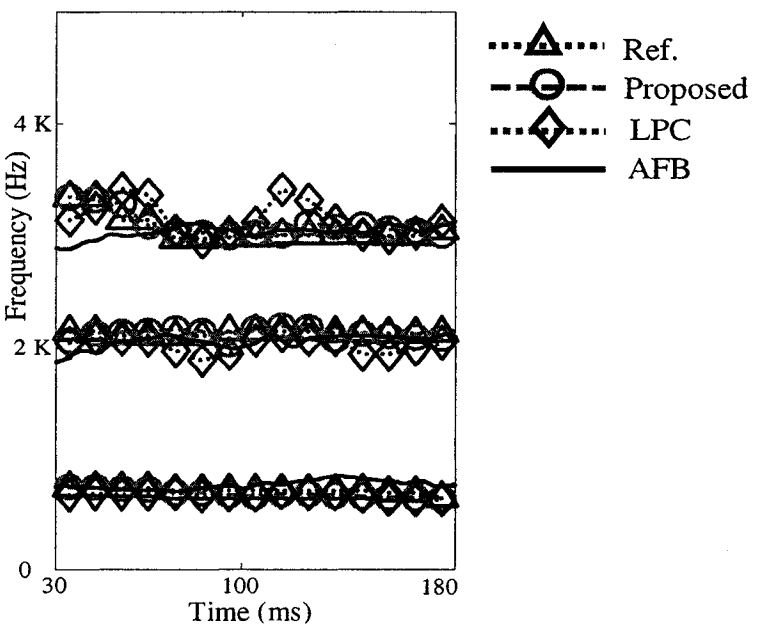
by different methods at SNR = 5 dB in the presence of white Gaussian noise are plotted on the spectrogram. In comparison to the estimation performance of the other methods, the proposed method can track the rise-fall pattern of different formants almost accurately.

6.7 Conclusion

In this Chapter, based on the proposed system identification techniques, formant frequency estimation methods have been presented. All of the proposed methods are capable of providing an accurate estimate of formant frequencies in the presence of noise. It is worth mentioning that in these methods, some of the previously proposed models have been effectively used jointly with the least-square model-fitting approach to extract the formant frequencies. The ARMA correlation model proposed in Chapter 2 has first been used to develop a formant estimation scheme. In a similar fashion, the AR ramp-cepstrum model that were proposed in Chapter 3 has then be used to develop another formant frequency estimation method. A ramp cosine cepstrum model for the once-repeated ACF of speech signals has been developed as a new principle, leading to a RCC model based formant frequency estimation method. The new method can effectively handle the noisy environment due to the



(a)



(b)

Figure 6.20: Formant tracks at SNR = 0 dB in the presence of white noise for (a) male vowel /a/ and (b) female vowel /e/.

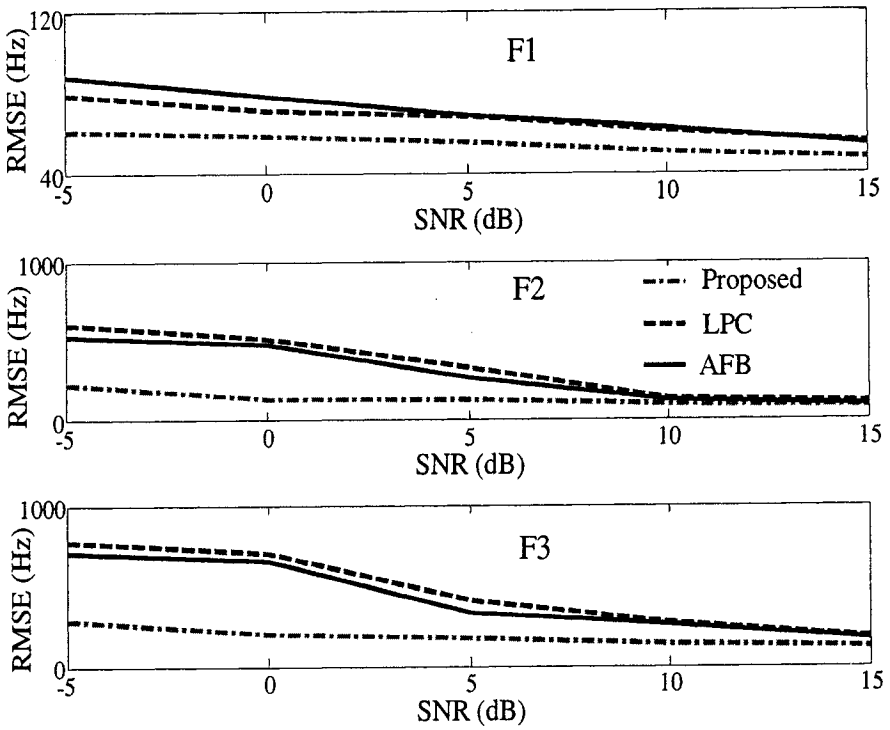
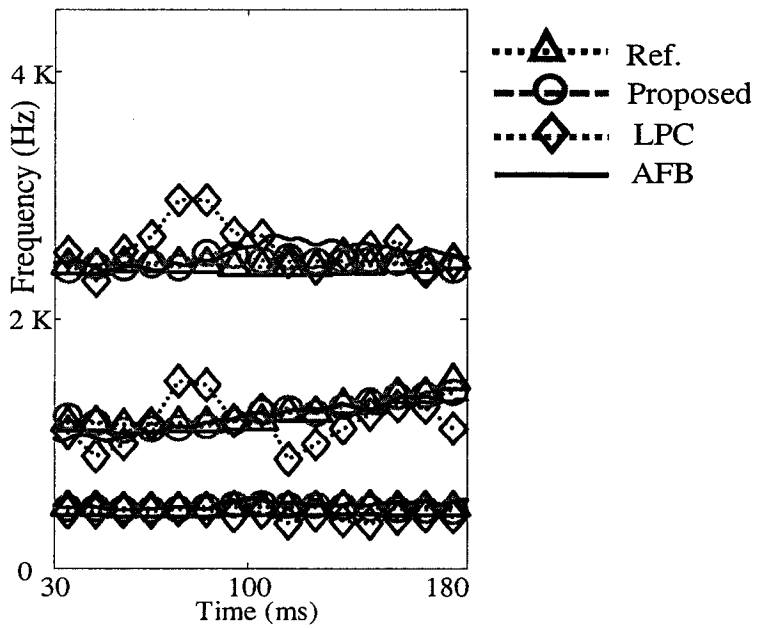
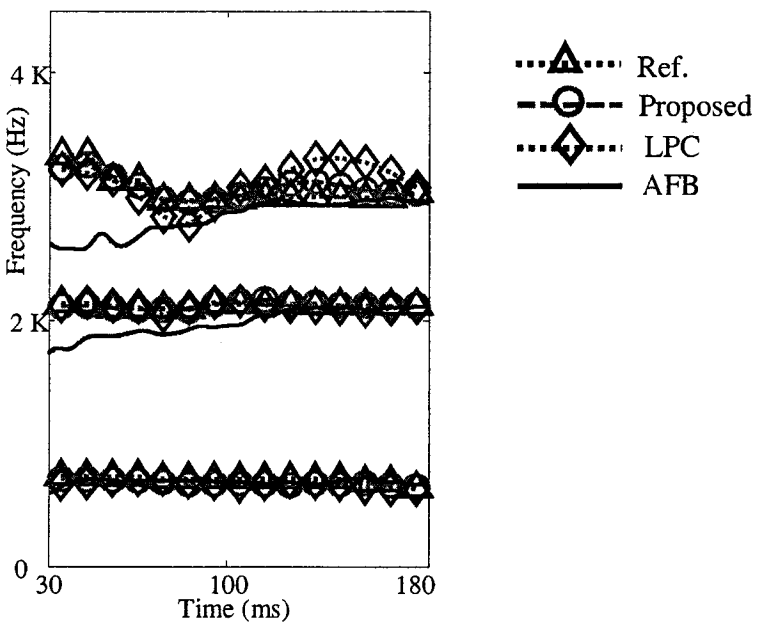


Figure 6.21: Effect of SNR on average RMSE (Hz) in the estimation of different formants (F1, F2, and F3).

combined advantages of the ramp-cepstrum model-fitting and the once-repeated correlation operations. An important feature of this method is its DCT/IDCT based implementation which offers several advantages including simple phase unwrapping. From extensive experimentations on synthetic and natural speech signals under noisy conditions, it has been found that the proposed method provides an accurate formant frequency estimate at a moderate to low levels of SNR.



(a)



(b)

Figure 6.22: Formant tracks at SNR = 0 dB in the presence of babble noise. (a) male vowel /u/ and (b) female vowel /e/.

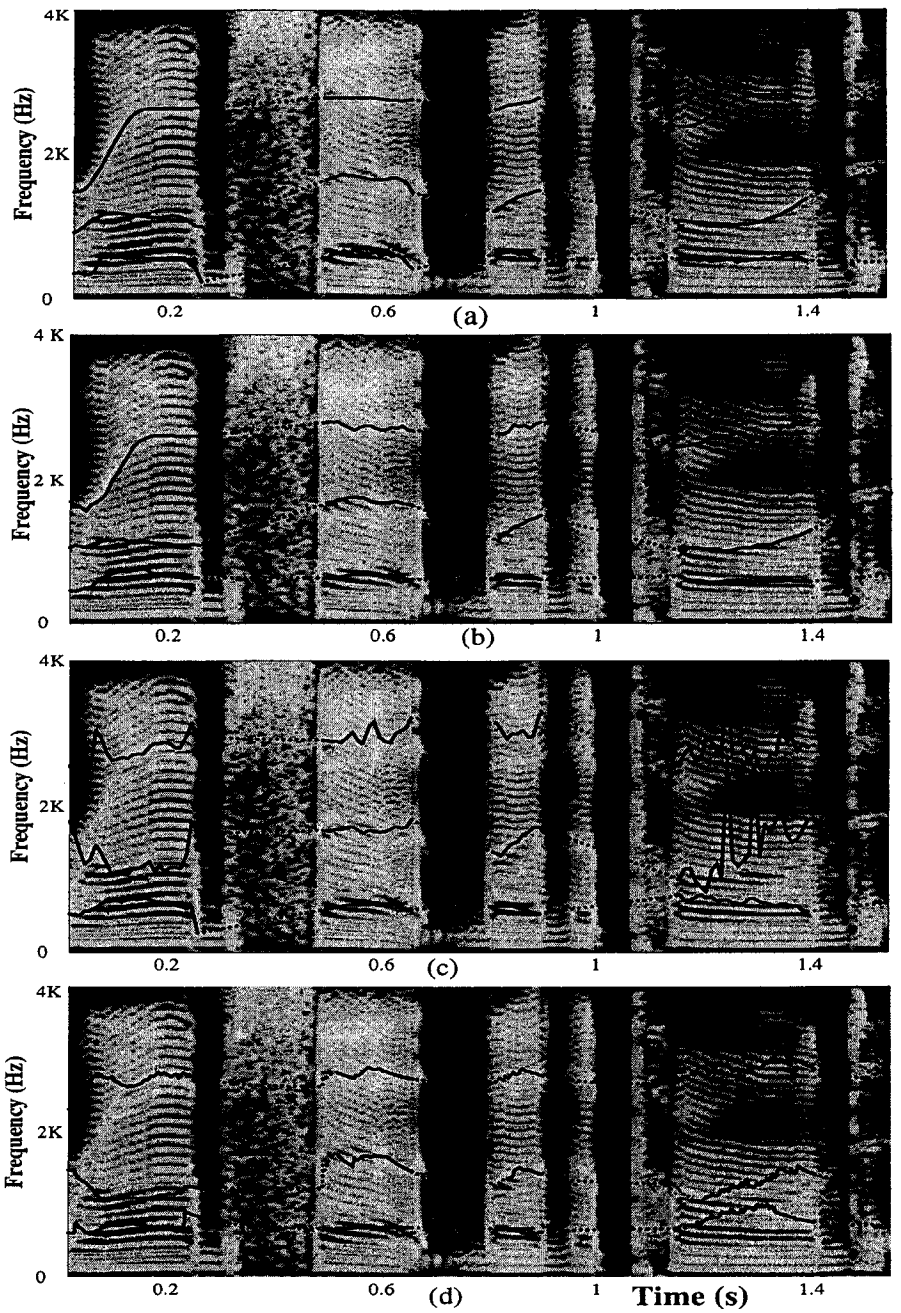


Figure 6.23: Formant estimation results for a male utterance "Rob sat by the pond" at SNR = 5 dB plotted on clean speech spectrogram; (a)Reference, (b) Proposed ramp cosine cepstrum model based method, (c) LPC method, and (d) AFB method.

Chapter 7

Conclusion

7.1 Concluding Remarks

Estimation of system parameters from given noisy observations has very important applications in various fields, such as signal processing, communication, and control. As far as real-life applications are concerned, parameter estimation using only the noise-corrupted observations at a very low SNR is a very difficult problem; yet there is a strong demand for noise-robust estimation methods. For example, formant estimation from noise-corrupted speech is essential in speech processing, but only a few research results are available in literature. In this dissertation, some effective methodologies for the identification of AR and ARMA systems, which are able to estimate accurately the parameters of AR and ARMA systems using the noise-corrupted observations at very low levels of SNR, have been developed.

In the development of the proposed system identification methods, the input to the system need not be accessible, and both white noise and periodic impulse-train inputs are taken into account in order to handle speech applications. A new correlation model for the ARMA signal has been proposed and then employed in a residue-based least-squares (RBLS) model-fitting optimization algorithm to estimate the AR

parameters of the ARMA system. To overcome the difficulty in identifying the MA part of the ARMA system under a heavy noisy condition, a new algorithm, which employs a noise-compensation in the ACF of the residual signal and then estimates the MA parameters from the noise-compensated ACF or from the inverse ACF, has been developed.

The ramp-cepstrum models of a one-sided autocorrelation function (OSACF) of the AR as well as ARMA signals have been proposed. The parameter estimation from noisy observations is then carried out by employing a ramp-cepstral model-fitting based approach. It has been shown that, once the AR parameters are obtained, the MA parameters can be successfully estimated using the MA estimation algorithm given in this thesis. Considering the implementation related advantages of the DCT, the ramp cosine cepstrum models of the OSACF of AR and ARMA signals have also been developed and these models are then used for the estimation of the system parameters through a process refer to as ramp cosine cepstral fitting.

In order to overcome the limitations of the conventional low-order Yule-Walker methods, a new noise-compensated quadratic eigenvalue method has been developed giving the estimate of the AR parameters of the ARMA system along with that of the noise variance. The MA parameters are then estimated by using the spectral factorization of a noise-compensated ACF of the residual signal.

As an application of the proposed system identification algorithms to speech analysis, some robust formant estimation methods have been developed which can efficiently handle the adverse effect of very noisy observations on the estimation. Based on the ARMA correlation, ramp-cepstrum and ramp cosine cepstrum models proposed in thesis, formant frequencies have been estimated from the observed noise-corrupted speech by employing an adaptive RBLS model-fitting algorithm. Estimation perfor-

mance of the proposed methods has been studied through extensive experimentations under heavy noisy conditions, and the results demonstrate the effectiveness in the system administration and superiority in terms of the estimation accuracy and consistency of the methods.

Some of the key contributes of the investigation undertaken in this thesis can be summarized as follows:

1. All the system identification methods have been developed with a target to obtain an accurate and consistent estimate of system parameters under heavy noisy conditions. The estimation accuracy of the system identification methods available in literatures, including those proposed to handle a noisy environment, deteriorates drastically with the increase of the level of noise. The identification methods proposed in this thesis provide a much superior performance even at very low levels of SNR.

2. Unlike many of the existing system identification methods, the proposed methods do not require an access to the input signal of the system. The parameters can be estimated accurately by considering the system excitations as a white Gaussian noise or periodic impulse-train.

3. The proposed identification techniques are readily applicable to both AR and ARMA systems.

4. Because of a better noise immunity of the new models, developed in the correlation and cepstral domains, the proposed methods by using these models in the least-squares model fitting framework provide a superior identification performance in the presence of heavy noise.

5. The development of the proposed models as well as the estimation algorithms using both white noise and periodic impulse-train excitations, which respectively correspond to the unvoiced and voiced speech sounds, make these models readily applicable to speech analysis. The relevance of this feature has been demonstrated for vocal-tract system identification and formant frequency estimation problems.

7.2 Scope for Further Work

The research work undertaken in this thesis can be extended in several aspects. One interesting area of investigation could be the handling of colored noise in the identification problem. In this research, we have carried out the identification process under a white noise observation of the system output. We have also successfully used the babble noise in the system identification as an illustration of the proposed techniques in a practical situation. A colored noise is very much application dependent and generally is not possible to have a single unified model for all the different types of colored noises. Usually, they are modeled as the output of a particular system driven by a white noise. One possible way is to identify the colored noise generating system. Then, an inverse filtering operation can be performed on noisy observations to whiten the noise and reduce the problem into a standard system identification task in the presence of white noise. The main problem will be the availability of noise information for its parametric identification. A further solution could be the application of an adaptive filter technique.

It has been shown that the proposed method can accurately extract formant frequencies, the most important speech feature, even under severe noisy conditions. This important feature of the proposed method creates room for its wide applications in the

various speech processing tasks, such as speech recognition and synthesis. Incorporation of formant features into a speech recognition system increases the recognition accuracy. Therefore, a noise-robust formant frequency estimator is in a great demand. In order to incorporate some additional acoustical information in the speech recognition system, formant tracking is sometimes helpful in addition to the specific formant values. Hence, a complete formant tracking system incorporating pitch detector, voicing detector, gender detector, and some logic blocks for proper tracking could be an interesting topic of future investigation.

Bibliography

[1] K. A. Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, no. 8, pp. 1310–1322, Aug. 1997.

[2] C.-Y. Chi, C.-Y. Chen, C.-H. Chen, and C.-C. Feng, "Batch processing algorithms for blind equalization using higher-order statistics," *IEEE Signal Processing Mag.*, vol. 20, no. 1, pp. 25–49, Jan. 2003.

[3] K. E. Baddour and N. C. Beaulieu, "Autoregressive modeling for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.

[4] J. R. Stack, T. G. Habetler, and R. G. Harley, "Bearing fault detection via autoregressive stator current modeling," *IEEE Trans. Ind. Applicat.*, vol. 40, no. 3, pp. 740–747, May. 2004.

[5] R. W. Wies, J. W. Pierre, and D. J. Trudnowski, "Use of ARMA block processing for estimating stationary low-frequency electromechanical modes of power systems," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 167–173, Feb. 2003.

[6] S. Hashimoto, K. Hara, H. Funato, and K. Kamiyama, "AR-based identification and control approach in vibration suppression," *IEEE Trans. Ind. Applicat.*, vol. 37, no. 3, pp. 806–811, May 2001.

[7] D. O'Shaughnessy, *Speech Communications Human and Machine*, 2nd ed. NY: IEEE Press, 2000.

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

[9] M. Yang, "Low bit rate speech coding," *IEEE Potentials*, vol. 23, no. 4, pp. 32–36, Oct. 2004.

[10] D. O'Shaughnessy, "Interacting with computers by voice: automatic speech recognition and synthesis," *Proc. IEEE*, vol. 91, no. 9, pp. 1272–1305, Sept. 2003.

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

[12] M. S. Zilovic, R. P. Ramachandran, and R. J. Mammone, "Speaker identification based on the use of robust cepstral features obtained from pole-zero transfer functions," *IEEE Trans. Speech Audio Processing*, vol. 6, no. 3, pp. 260–267, May 1998.

[13] J. Alvarez, M. Ramon, M. P. Cumplido, A. A. Rodriguez, and A. R. F. Vidal, "Support vector method for robust ARMA system identification," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 155–164, Jan. 2004.

[14] S. Lu, K. H. Ju, and K. H. Chon, "A new algorithm for linear and nonlinear ARMA model parameter estimation using affine geometry," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 10, pp. 1116–1124, Oct. 2001.

[15] M. Khadra, A. S. Al-Fahoum, and S. Binajjaj, "Quantitative analysis approach for cardiac arrhythmia classification using higher order spectral techniques," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 11, pp. 1840–1845, Nov. 2005.

[16] A. Nair, D. Calvetti, and D. G. Vince, "Regularized autoregressive analysis of intravascular ultrasound backscatter: improvement in spatial accuracy of tissue maps," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 51, no. 4, pp. 420–431, Jan. 2004.

[17] A. Vizinho and L. R. Wyatt, "Evaluation of the use of the modified-covariance method in HF radar ocean measurement," *IEEE J. Oceanic Eng.*, vol. 26, no. 4, pp. 832–840, Oct. 2001.

[18] K.-T. Kim, J.-H. Bae, and H.-T. Kim, "Effect of AR model-based data extrapolation on target recognition performance," *IEEE Trans. Antennas Propagat.*, vol. 51, no. 4, pp. 912–914, Jan. 2003.

[19] J.-L. Zarader, A. Dabas, P. H. Flamant, B. Gas, and O. Adam, "Adaptive parametric algorithms for processing coherent doppler-lidar signal," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, no. 6, pp. 2678–2691, Nov. 1999.

[20] O. Rosec, J.-M. Boucher, B. Nsiri, and T. Chonavel, "Blind marine seismic deconvolution using statistical MCMC methods," *IEEE J. Oceanic Eng.*, vol. 28, no. 3, pp. 502–512, Jul. 2003.

[21] G. I. Bourdopoulos and T. L. Deliyannis, "High-order vector sigma-delta modulators," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 6, pp. 493–503, June 2000.

[22] Y. Fang and T. W. S. Chow, "Orthogonal wavelet neural networks applying to identification of Wiener model," *IEEE Trans. Circuits Syst. I*, vol. 47, no. 4, pp. 591–593, April 2000.

[23] D. Veselinovic and D. Graupe, "A wavelet transform approach to blind adaptive filtering of speech from unknown noises," *IEEE Trans. Circuits Syst. II*, vol. 50, no. 3, pp. 150–154, Mar. 2003.

[24] Q. Tian, B.-S. Sun, A. Ozguler, S. A. Morris, and W. D. OBrien, "Parametric modeling in food package defect imaging," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 47, no. 3, pp. 635–643, May 2000.

[25] S. M. Kay, "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, no. 1, pp. 56–65, Feb. 1985.

[26] L. T. McWhorter and L. L. Scharf, "Nonlinear maximum likelihood estimation of autoregressive time series," *IEEE Trans. Signal Processing*, vol. 43, no. 12, pp. 2909–2919, Dec. 1995.

[27] M. L. Vis and L. L. Scharf, "A note on recursive maximum likelihood for autoregressive modeling," *IEEE Trans. Signal Processing*, vol. 42, no. 10, pp. 2881–2883, Oct. 1994.

[28] S. Degerine, "A necessary and sufficient condition for the existence of the maximum likelihood estimate in autoregressive models," *IEEE Trans. Signal Processing*, vol. 41, no. 2, pp. 988–990, Feb. 1993.

[29] B. Armour and S. D. Morgera, "An exact forward-backward maximum likelihood autoregressive parameter estimation method," *IEEE Trans. Signal Processing*, vol. 39, no. 9, pp. 1985–1993, Sept. 1991.

[30] B. C. Frias and J. D. Rogers, "On the exact maximum likelihood estimation of Gaussian autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 6, pp. 922–924, Jun. 1988.

[31] C. F. Ansley, "An algorithm for the exact likelihood of a mixed autoregressive-moving average process," *Biometrika*, vol. 66, no. 1, pp. 59–65, Apr. 1979.

[32] H. Akaike, "Maximum likelihood identification of Gaussian autoregressive moving average models," *Biometrika*, vol. 60, no. 2, pp. 255–265, Aug. 1973.

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

[34] L. Ljung, *System Identification: Theory for the User.* Upper Saddle River, NJ: Prentice-Hall Ltd., 1999.

[35] N. Xie and H. Leung, "Blind identification of autoregressive system using chaos," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 9, pp. 1953–1964, Sept. 2005.

[36] H. Leung, "System identification using chaos with application to equalization of a chaotic modulation system," *IEEE Trans. Circuits Syst. I*, vol. 45, no. 3, pp. 314–320, Mar. 1998.

[37] H. Leung, S. Wang, and A. M. Chan, "Blind identification of an autoregressive system using a nonlinear dynamical approach," *IEEE Trans. Signal Processing*, vol. 48, no. 1, pp. 3017–3027, Nov. 2000.

[38] M. Kaveh, "High resolution spectral estimation for noisy signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 27, no. 3, pp. 286–287, June 1979.

[39] J. F. Kinkel, J. Perl, L. Scharf, and A. Stubberud, "A note on covariance-invariant digital filter design and autoregressive-moving average spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 27, no. 2, pp. 200–202, Apr. 1979.

[40] S. M. Kay, "A new ARMA spectral estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, no. 5, pp. 585–588, Oct. 1980.

[41] J. A. Cadzow, "High performance spectral estimation-a new ARMA method," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, no. 5, pp. 524–529, Oct. 1980.

[42] B. Friedlander and B. Porat, "A spectral matching technique for ARMA parameter estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 2, pp. 338–343, Apr. 1984.

[43] X.-D. Zhang and H. Takeda, "An approach to time series analysis and ARMA spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 9, pp. 1303–1313, Sept. 1987.

[44] R. Moses and A. A. Beex, "A comparison of numerator estimators for ARMA spectra," *EEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 6, pp. 1668–1671, Dec. 1986.

[45] A. Kaderli and A. S. Kayhan, "Spectral estimation of ARMA processes using ARMA-cepstrum recursion," *IEEE Signal Process. Lett.*, vol. 7, no. 9, pp. 259–261, Sept. 2000.

[46] P. Stoica, T. Soderstrom, and B. Friedlander, "Optimal instrumental variable estimates of the AR parameters of an ARMA process," *IEEE Trans. Automat. Contr.*, vol. 30, no. 11, pp. 1066–1074, Nov. 1985.

[47] N. Zhou and J. W. Pierre, "Estimation of autoregressive parameters by the constrained total least square algorithm using a bootstrap method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, Philadelphia, PA, Mar. 2005, pp. 417–420.

[48] D. P. Burke, S. P. Kelly, P. D. Chazal, R. B. Reilly, and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *IEEE Trans. Neural Syst. Rehab. Eng.*, vol. 13, no. 1, pp. 12–17, Mar. 2005.

[49] K. K. Paliwal, "Estimation of noise variance from the noisy AR signal and its application in speech enhancement," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 2, pp. 292–294, Feb. 1988.

[50] S. M. Kay, "Noise compensation for autoregressive spectral estimates," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, no. 3, pp. 292–303, Jun. 1980.

[51] ——, "The effects of noise on the autoregressive spectral estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 27, no. 5, pp. 478–485, Oct. 1980.

[52] Y. Chan and R. Langford, "Spectral estimation via the high-order Yule-Walker equations," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 30, no. 5, pp. 689–698, Jun. 1982.

[53] D. F. Gingras and E. Masry, "Autoregressive spectral estimation in additive noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 490–501, Apr. 1988.

[54] L. V. Dominguez, "New insights into the higher-order Yule-Walker equations," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 9, pp. 1649–1651, Sept. 1990.

[55] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, no. 9, pp. 905–939, Sept. 1982.

[56] B. Porat and B. Friedlander, "Asymptotic analysis of the bias of the modified Yule-Walker estimator," *IEEE Trans. Automat. Contr.*, vol. 30, no. 8, pp. 765–767, Aug. 1985.

[57] B. Friedlander and K. C. Sharman, "Performance evaluation of the modified Yule-Walker estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, no. 3, pp. 719–725, Jun. 1985.

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

[59] ——, "On the noise-compensated Yule-Walker equations," *IEEE Trans. Signal Processing*, vol. 49, no. 6, pp. 1119–1121, Jun. 2001.

[60] W.-R. Wu and P.-C. Chen, "Adaptive AR modeling in white Gaussian noise," *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1184–1192, May. 1997.

[61] W. X. Zheng, "An efficient method for estimation of autoregressive signals in noise," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Kobe, Japan, May 2005, pp. 1433–1436.

[62] ——, "Fast identification of autoregressive signals from noisy observations," *IEEE Trans. Circuits Syst. II*, vol. 52, no. 1, pp. 43–48, Jan. 2005.

[63] ——, "Autoregressive parameter estimation from noisy data," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 1, pp. 71–75, Jan. 2000.

[64] ——, "A least-squares based method for autoregressive signals in the presence of noise," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 1, pp. 81–85, Jan. 1999.

[65] C.-Y. Chi, C.-C. Feng, C.-H. Chen, and C.-Y. Chen, *Blind Equalization and System Identification: Batch Processing Algorithms, Performance and Applications.* london, UK: Springer-Verlag, 2006.

[66] H. Kanai, M. Abe, and K. Kido, "Accurate autoregressive spectrum estimation at low signal-to-noise ratio using a phase matching technique," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 9, pp. 1264–1272, Sept. 1987.

[67] F. Ferdousi, A. T. Connie, M. Sharmin, and M. R. Khan, "System identification at an extremely low SNR using energy density in DCT domain," *IEEE Signal Process. Lett.*, vol. 12, no. 4, pp. 289–292, Apr. 2005.

[68] P. Stoica, T. McKelvey, and J. Mari, "MA estimation in polynomial time," *IEEE Trans. Signal Processing*, vol. 48, no. 7, pp. 1999–2012, July 2000.

[69] A. V. Oppenheim and R. W. Schafer, "From frequency to quefrency: a history of the cepstrum," *IEEE Signal Processing Mag.*, vol. 21, no. 5, pp. 95–106, Sept. 2004.

[70] L. Deng, J. Droppo, and A. Acero, "Estimating cepstrum of speech under the presence of noise using a joint prior of static and dynamic features," *IEEE Signal Processing Mag.*, vol. 12, no. 3, pp. 218–233, May 2004.

[71] 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 Trans. Speech Audio Processing*, vol. 11, no. 5, pp. 435–446, Sept. 2003.

[72] J. Xu, J. Cheng, and Y. Wu, "A cepstral method for analysis of acoustic transmission characteristics of respiratory system," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 5, pp. 660–664, May 1998.

[73] Z. Huang, X. Yang, X. Zhu, and A. Kuh, "Homomorphic linear predictive coding: a new estimation algorithm for all-pole speech modeling," *IEE Proc. Comm. Speech Vis.*, vol. 137, no. 2, pp. 103–108, April 1990.

[74] M. A. L. Hernandez, P. Stoica, and M. A. Rojas, "ARMA parameter estimation: Revisiting a cepstrum-based method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 3685–3688.

[75] S. Li and B. W. Dickinson, "Application of the lattice filter to robust estimation of AR and ARMA models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 4, pp. 502–512, Apr. 1988.

[76] N. M. Hossain, M. S. I. Patwary, A. M. M. Reza, and M. K. Hasan, "An effective autocorrelation model for estimation of ARMA parameters from noisy observations," in *Proc. Int. Work. Systems, Signals Image Process.*, Poznan, Poland, Sept. 2004, pp. 295–298.

[77] M. K. Hasan, N. M. Hossain, and P. A. Naylor, "A novel autocorrelation model-based identification method for ARMA systems in noise," *IEE Proc. Vis. Image Signal Process.*, vol. 152, no. 5, pp. 520–526, Oct. 2005.

[78] D. He and H. Leung, "Semi-blind identification of ARMA systems using a dynamic-based approach," *IEEE Trans. Circuits Syst. I*, vol. 52, no. 1, pp. 179–190, Jan. 2005.

[79] K. Lo, H. Kimura, W.-H. Kwon, and X. Yang, "Empirical frequency-domain optimal parameter estimate for black-box processes," *IEEE Trans. Circuits Syst. I*, vol. 53, no. 2, pp. 419–430, Feb. 2006.

[80] P. Stoica, H. Li, and L. Jian, "Amplitude estimation of sinusoidal signals: survey, new results, and an application," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 338–352, Feb. 2000.

[81] Haskins Laboratory. Articulatory synthesys. [Online]. Available: http://www.haskins.yale.edu/facilities/asy.html.

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

[83] K. Mustafa and I. C. Bruce, "Robust formant tracking for continuous speech with speaker variability," *IEEE Trans. Audio, Speech Language Processing*, vol. 14, no. 2, pp. 435–444, Mar. 2006.

[84] R. C. Snell and F. Milinazzo, "Formant location from LPC analysis data," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 2, pp. 129–134, Apr. 1993.

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

[86] W. Verhelst and O. Steenhaut, "A new model for the short-time complex cepstrum of voiced speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 1, pp. 43–51, Feb. 1986.

[87] L. Deng, A. Acero, and I. Bazzi, "Tracking vocal tract resonances using a quantized nonlinear function embedded in a temporal constraint," *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 2, pp. 425–434, Mar. 2006.

[88] A. Watanabe, "Formant estimation method using inverse-filter control," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 4, pp. 317–326, May 2001.

[89] J. Malkin, X. Li, and J. Bilmes, "A graphical model for formant tracking," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 1, Philadelphia, PA, Mar. 2005, pp. 913–916.

[90] Y. Shi and E. Chang, "Spectrogram-based formant tracking via particle filters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 1, Hong Kong, Apr. 2003, pp. 168–171.

[91] Y. Zheng and M. H. Johnson, "Formant tracking by mixture state particle filter," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 1, Montreal, Canada, Mar. 2004, pp. 565–568.

[92] D. J. Nelson, "Cross-spectral based formant estimation and alignment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing (ICASSP)*, vol. 2, Montreal, Canada, Mar. 2004, pp. 621–624.

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

[94] B. Chen and P. C. Loizou, "Formant frequency estimation in noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 1, Montreal, Canada, May 2004, pp. 581–584.

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

[96] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "A novel technique for the identification of an ARMA system under very low levels of SNR," *IEEE Trans. Circuits Syst. I*, vol. 55, no. 7, pp. 1988–2001, Aug. 2008.

[97] ——, "An approach to ARMA system identification at a very low signal-to-noise ratio," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, Philadelphia, PA, Mar. 2005, pp. 259–261.

[98] P. Stoica, "Model-order selection: a review of information criterion rules," *IEEE Signal Processing Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[99] P. M. T. Broersen and S. D. Waele, "Finite sample properties of ARMA order selection," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 3, pp. 645–651, June 2004.

[100] 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 Trans. Circuits Syst. I*, vol. 52, no. 7, pp. 1394–1405, July 2005.

[101] P. Stoica, T. McKelvey, and J. Mari, "MA estimation in polynomial time," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1999–2012, July 2000.

[102] C. Chatfield, "Inverse autocorrelations," *J. Roy. Statist. Soc. Ser. A*, vol. 142, no. 3, pp. 363–377, 1979.

[103] C. Shahnaz, W.-P. Zhu, and M. O. Ahmad, "Robust pitch estimation at very low snr exploiting time and frequency domain cues," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, (ICASSP)*, vol. 1, Philadelphia, PA, Mar. 2005, pp. 389–392.

[104] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Identification of autoregressive systems in noise based on a ramp cepstrum model," *IEEE Transaction on Circuits and Systems II*, 2008, in press, available on line in the IEEE Xplore; Digital Object Identifier: 10.1109/TCSII.2008.925660.

[105] ——, "Identification of ARMA systems in noise based on a ramp cepstrum model," *submitted to IET Signal Process.*

[106] ——, "Noisy autoregressive system identification by the ramp cepstrum of one-sided autocorrelation function," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Kobe, Japan, May 2005, pp. 3147–3150.

[107] S. A. Fattah, W. P. Zhu, and M. O. Ahmad, "A blind ARMA system identification technique in noise," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Island of Kos, Greece, May 2006, pp. 21–24.

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

[109] L. Weruaga, "All-pole estimation in spectral domain," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4821–4830, Oct. 2007.

[110] P. M. T. Broersen, "Autoregressive model orders for durbin's MA and ARMA estimators," *IEEE Trans. Signal Processing*, vol. 48, no. 8, pp. 2454–2457, Aug. 2000.

[111] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "A ramp cosine cepstrum model for the parameter estimation of autoregressive signals with low SNR," *submitted to IEEE Trans. Signal Process.*

[112] ——, "A ramp cosine cepstrum model for the parameter estimation of autoregressive moving average signals with low SNR," *in preparation*, to be submitted shortly.

[113] F. Wang and P. Yip, "Cepstrum analysis using discrete trigonometric transform," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 538–541, Feb. 1991.

[114] B. Bhanu and J. H. McClellan, "On the computation of the complex cepstrum," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 28, no. 5, pp. 583–585, Oct. 1980.

[115] D. G. Long, "Exact computation of the unwrapped phase of a finite length time series," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 11, pp. 1787–1790, Nov. 1988.

[116] G. E. Kopec, A. V. Oppenheim, and J. M. Tribolet, "Speech analysis by homomorphic prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 25, no. 1, pp. 40–49, Feb. 1977.

[117] T. Kobayashi and S. Imai, "Spectrum analysis using generalized cepstrum," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 5, pp. 1087–1089, Oct. 1984.

[118] K. Steiglitz, "On the simultaneous estimation of poles and zeros in speech analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 25, no. 3, pp. 229–234, June 1977.

[119] 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 Proc. Vis. Image Signal Process.*, vol. 145, no. 5, pp. 316–321, Oct. 1998.

[120] J. Garofolo, L. L. W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," in *Proc. Ling. Data Consort.*, 1993.

[121] J. M. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3099–3111, May 1995.

[122] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 12, no. 3, p. 247251, Jul. 1993.

[123] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Identification of ARMA systems based on noise-compensation in the correlation domain," *submitted to IEEE Trans. Circuits Syst. I.*

[124] S. A. Fattah, W. P. Zhu, and M. O. Ahmad, "An algorithm for ARMA model parameter estimation from noisy observations," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Seattle,WA, May 2008, pp. 3202–3205.

[125] ——, "Identification of autoregressive moving average systems from noise-corrupted observations," in *Proc. IEEE Int. Northeast Workshop Circuits Syst. (NEWCAS)*, Montreal, Canada, Jun. 2008, pp. 69–72.

[126] ——, "A correlation domain algorithm for autoregressive system identification from noisy observations," in *Proc. IEEE Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Knoxville, TN, Aug. 2008, pp. 934–937.

[127] ——, "An algorithm for the identification of autoregressive moving average systems from noisy observations," in *Proc. IEEE Canadian Conf. on Elect. and Comp. Eng. (CCECE)*, Niagara Falls,Canada, May 2008, pp. 1815–1818.

[128] ——, "An identification technique for noisy ARMA systems in correlation domain," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, New Orleans,LA, May 2007, pp. 349–352.

[129] P. Stoica and R. Moses, *Spectral Analysis of Signals.* Upper Saddle River, NJ: Prentice-Hall, 2005.

[130] M. Elad, P. Milanfar, and G. H. Golub, "Shape from moments- an estimation theory perspective," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1814–1829, July 2004.

[131] M. S. Misrikhanov and V. N. Ryabchenko, "The quadratic eigenvalue problem in electric power systems," *Automation Remote Control*, vol. 67, no. 5, pp. 698–720, 2006.

[132] G. H. Golub and C. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[133] P. M. T. Broersen and S. D. Waele, "Automatic identification of time-series models from long autoregressive models," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 5, pp. 1862–1868, Oct. 2005.

[134] S. A. Fattah, W. P. Zhu, and M. O. Ahmad, "An approach to formant frequency estimation at a very low signal-to-noise ratio," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, Honolulu, HI, Apr. 2007, pp. 469–472.

[135] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "An algorithm for formant frequency estimation from noise-corrupted speech signals," *Published in Canadian*

Acoustics, vol. 35, no. 3, pp. 110–111, Oct. 2007, presented in the annual conf. of the Canadian Acoustical Association(CAA).

[136] ——, "Formant frequency estimation of speech signals at low SNR based on a correlation model-fitting algorithm," *in preparation*, to be submitted shortly.

[137] S. A. Fattah, W. P. Zhu, and M. O. Ahmad, "A formant frequency estimation scheme for speech signals in the presence of noise," in *Proc. IEEE Int. Symp. Signals, Systems Elect. (ISSSE)*, Montreal, Canada, Aug. 2007, pp. 407–410, (Received URSI Canadian Young Scientist Award 2007 for this paper).

[138] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "Development of a speaker independent vowel recognition system," *Presented in the centre for advanced systems and communications (SYTACom) research workshop, Quebec City, Canada*, May 2007.

[139] S. A. Fattah, W. P. Zhu, and M. O. Ahmad, "A formant frequency estimation algorithm for speech signals with low signal-to-noise ratio," in *Proc. IEEE Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Montreal, Canada, Aug. 2007, pp. 105–108.

[140] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "A formant frequency estimator for noisy speech based on correlation and cepstrum," *to be published in Canadian Acoustics*, Oct. 2008, to be presented in the annual conf. of the Canadian Acoustical Association(CAA).

[141] ——, "Formant frequency estimation of speech signals at low SNR based on a ramp cepstral model-fitting algorithm," *in preparation*, to be submitted shortly.

[142] ——, "An algorithm for formant feature extraction from noise-corrupted speech," *Presented in the centre for advanced systems and communications (SYTACom) research poster session, Montreal, Canada*, Jul. 2008.

[143] S. A. Fattah, W. P. Zhu, and M. O. Ahmad, "A cepstral domain algorithm for formant frequency estimation from noise-corrupted speech," in *Proc. IEEE Int. Conf. Neural Networks, Signal Process. (ICNNSP)*, Zhenjiang, China, Jun. 2008, pp. 114–119, (Nominated for the best paper award).

[144] S. A. Fattah, W.-P. Zhu, and M. O. Ahmad, "A formant estimation scheme for speech signals under noisy conditions," *Presented in the centre for advanced systems and communications (SYTACom) research workshop, Montreal, Canada*, May 2008, (First prize winner in the SYTACom research poster competition).

[145] ——, "Formant frequency estimation of speech signals at low SNR based on a DCT ramp cepstrum of repeated autocorrelation function," *submitted to IEEE Trans. Audio, Speech Language Process.*

[146] H. Morikawa and H. Fujisaki, "System identification of the speech production process based on a state-space representation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, no. 2, pp. 252–262, Apr. 1984.

[147] L. Deng, X. Cui, R. Pruvenok, J. Huang, S. Momen, Y. Chen, and A. Alwan, "A database of vocal tract resonance trajectories for research in speech processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 1, Toulouse, France, Apr. 2006, pp. 369–372.

[148] S. A. Fattah, W. P. Zhu, and M. O. Ahmad, "An ARMA system identification scheme in the presence of noise," in *Proc. IEEE Int. Northeast Workshop Circuits Syst. (NEWCAS)*, Montreal, Canada, Aug. 2007, pp. 602–605.

[149] ——, "An identification technique for ARMA systems in the presence of noise," in *Proc. IEEE Canadian Conf. on Elect. and Comp. Eng. (CCECE)*, Vancouver, Canada, Apr. 2007, pp. 888–891.

[150] ——, "Noisy autoregressive system identification based on repeated autocorrelation function," in *Proc. IEEE Canadian Conf. on Elect. and Comp. Eng. (CCECE)*, Ottawa, Canada, May 2006, pp. 1572–1575.

Appendix A

Derivation of the ARMAC Model for the White Noise Excitation

Equation (2.10) can be rewritten easily as

$$r_x(\tau) = \sum_{k=1}^{P}\sum_{j=1}^{P} \eta_k \eta_j \left(I_1 + I_2\right); \tau = 0, 1, \ldots, M-1; \tag{A1}$$

$$I_1 = \frac{1}{N}\sum_{n=0}^{\Omega}\sum_{l=0}^{n} u(l)^2 p_k^{\tau+n-l} p_j^{n-l}; \tag{A2}$$

$$I_2 = \frac{1}{N}\sum_{n=0}^{\Omega}\sum_{l=0}^{n}\sum_{m=0, m\neq l}^{n+\tau} u(l)u(m) p_k^{\tau+n-m} p_j^{n-l} \tag{A3}$$

with $\Omega = N - \tau - 1$. By regrouping the terms in (A3), it can be shown that I_2 depends on the non-zero lags of the time-domain autocorrelation of the white noise $u(n)$, which can be neglected. Expanding the first summation, the expression for I_1 given by (A2) can be rewritten as

$$\begin{aligned} I_1 = {} & \frac{1}{N}u(0)^2 p_k^{\tau} + \frac{1}{N}\sum_{l=0}^{1} u(l)^2 p_k^{\tau+1-l} p_j^{1-l} + \frac{1}{N}\sum_{l=0}^{2} u(l)^2 p_k^{\tau+2-l} p_j^{2-l} + \cdots \\ & + \frac{1}{N}\sum_{l=0}^{\Omega} u(l)^2 p_k^{\tau+\Omega-l} p_j^{\Omega-l} \end{aligned} \tag{A4}$$

Further expanding the summations in (A4), I_1 can be written as

$$\begin{aligned} I_1 &= \frac{1}{N}[u(0)^2 p_k^\tau p_j^0] + \frac{1}{N}[u(0)^2 p_k^{\tau+1} p_j^1 + u(1)^2 p_k^\tau p_j^0] \\ &+ \frac{1}{N}[u(0)^2 p_k^{\tau+2} p_j^2 + u(1)^2 p_k^{\tau+1} p_j^1 + u(2)^2 p_k^\tau p_j^0] + \dots \\ &+ \frac{1}{N}[u(0)^2 p_k^{\tau+\Omega} p_j^\Omega + u(1)^2 p_k^{\tau+\Omega-1} p_j^{\Omega-1} + \dots + u(\Omega)^2 p_k^\tau p_j^0] \end{aligned} \tag{A5}$$

The above expression for I_1 can be rearranged as

$$\begin{aligned} I_1 &= \frac{1}{N} p_k^\tau p_j^0 [u(0)^2 + u(1)^2 + \dots + u(\Omega-1)^2 + u(\Omega)^2] \\ &+ \frac{1}{N} p_k^{\tau+1} p_j^1 [u(0)^2 + u(1)^2 + \dots + u(\Omega-1)^2] \\ &\quad \vdots \quad \vdots \\ &+ \frac{1}{N} p_k^{\tau+\Omega-1} p_j^{\Omega-1} [u(0)^2 + u(1)^2] \\ &+ \frac{1}{N} p_k^{\tau+\Omega} p_j^\Omega [u(0)^2] \end{aligned} \tag{A6}$$

Thus, I_1 can be expressed in a compact form as given by

$$I_1 = \frac{1}{N} \sum_{i=0}^{\Omega} p_k^{\tau+i} p_j^i \left[\sum_{l=0}^{\Omega-i} u(l)^2 \right] \tag{A7}$$

With increasing i, the terms in the first sum of the above equation, which correspond to higher powers of the poles decrease rapidly. Accordingly, it is sufficient to retain only the first φ terms in the first summation of (A7), where $\varphi \ll \Omega$ and since in the range $i = 0, 1, \dots, \varphi$, $\Omega - i \approx \Omega$ for the inner sum within the square brackets. Therefore, (A7) can be simplified as

$$I_1 = \chi \sum_{i=0}^{\varphi} p_k^{\tau+i} p_j^i = \chi \frac{1 - (p_k p_j)^{\varphi+1}}{1 - p_k p_j} p_k^\tau \tag{A8}$$

where $\chi = \frac{1}{N} \left[\sum_{l=0}^{\Omega} u(l)^2 \right]$. For a reasonably large value of φ, $(p_k p_j)^{\varphi+1}$ can be neglected and (A8) reduces to (2.11).

Appendix B

Derivation of the ARMAC Model for the Periodic Impulse-train Excitation

Consider first $r_x(0)$ which can be expanded using (2.5) as

$$r_x(0) = \left[r_{x_1}(0) + \cdots + r_{x_{(\lambda-1)}}(0) + r_{x_R}(0)\right](1/N); \tag{B1}$$

where

$$r_{x_1}(0) = x^2(0) + \cdots + x^2(T-1) = \sum_{k=1}^{P}\sum_{j=1}^{P} \eta_k \eta_j \frac{1 - p_k^T p_j^T}{1 - p_k p_j}$$

$$\vdots$$

$$r_{x_{(\lambda-1)}}(0) = \sum_{k=1}^{P}\sum_{j=1}^{P} \eta_k \eta_j \left(\frac{1 - p_k^T p_j^T}{1 - p_k p_j}\right) \times \left(\frac{1 - p_k^{(\lambda-1)T}}{1 - p_k^T}\right)\left(\frac{1 - p_j^{(\lambda-1)T}}{1 - p_j^T}\right)$$

$$\begin{aligned} r_{x_R}(0) &= x\{(\lambda-1)T\}^2 + \cdots + x\{(\lambda-1)T + R - 1\}^2 \\ &= \sum_{k=1}^{P}\sum_{j=1}^{P} \eta_k \eta_j \left(\frac{1 - p_k^R p_j^R}{1 - p_k p_j}\right) \times \left(\frac{1 - p_k^{\lambda T}}{1 - p_k^T}\right)\left(\frac{1 - p_j^{\lambda T}}{1 - p_j^T}\right) \end{aligned}$$

Using the expression for $r_{x_1}(0)$, $r_{x_2}(0)$ etc. as obtained above, (B1) can be expressed as

$$r_x(0) = \frac{1}{N}\sum_{k=1}^{P}\sum_{j=1}^{P}\eta_k\eta_j p_k^0\Theta_I, \tag{B2}$$

$$\Theta_I = \left[\frac{1-p_k^T p_j^T}{(1-p_k p_j)}\right][(\lambda-1)-g(p_k)-g(p_j)+g(p_k p_j)] + \left(\frac{1-p_k^R p_j^R}{1-p_k p_j}\right)\left(\frac{1-p_k^{\lambda T}}{1-p_k^T}\right)\left(\frac{1-p_j^{\lambda T}}{1-p_j^T}\right),$$

$$g(\alpha) = \alpha^T(1-\alpha^{(\lambda-1)T})/(1-\alpha^T) \tag{B3}$$

Similarly, $r_x(1)$ can be expanded as

$$r_x(1) = \left[r_{x_1}(1)+\cdots+r_{x_{(\lambda-1)}}(1)+r_{x_R}(1)\right](1/N) \tag{B4}$$

$$\begin{aligned} r_{x_1}(1) &= x(0)x(1)+\cdots+x(T-1)x(T) \\ &= \sum_{k=1}^{P}\sum_{j=1}^{P}\eta_k\eta_j p_k^1\left[\frac{1-p_k^T p_j^T}{1-p_k p_j}+p_j^T(p_k p_j)^{-1}\right] \end{aligned}$$

$$\vdots$$

$$\begin{aligned} r_{x_{(\lambda-1)}}(1) &= x\{(\lambda-2)T\}x\{(\lambda-2)T+1\}+\cdots \\ &\quad +x\{(\lambda-1)T-1\}x\{(\lambda-1)T\} \\ &= \sum_{k=1}^{P}\sum_{j=1}^{P}\eta_k\eta_j p_k^1(g_1+g_2), \end{aligned}$$

$$\begin{aligned} g_1 &= \left[\frac{1-(p_k p_j)^T}{1-p_k p_j}\right]\left[\frac{1-p_k^{(\lambda-1)T}}{1-p_k^T}\right]\left[\frac{1-p_j^{(\lambda-1)T}}{1-p_j^T}\right], \\ g_2 &= p_j^T(p_k p_j)^{-1}(1-p_j^{(\lambda-1)T})/(1-p_j^T), \end{aligned}$$

$$
\begin{aligned}
r_{x_R}(1) &= x\{(\lambda-1)T\}x\{(\lambda-1)T+1\}+\cdots \\
&+ x\{(\lambda-1)T+R-2\}x\{(\lambda-1)T+R-1\} \\
&= \sum_{k=1}^{P}\sum_{j=1}^{P}\eta_k\eta_j p_k^1\left[\frac{1-p_k^{\lambda T}}{1-p_k^T}\right]\left[\frac{1-p_j^{\lambda T}}{1-p_j^T}\right] \\
&\quad\times\left[\frac{1-(p_kp_j)^{R-1}}{1-p_kp_j}\right]
\end{aligned}
$$

It is to be noted in $r_x(1)$, the product terms containing two different expressions of G_{kn}, such as $x(T-1)x(T)$ in $r_{x_1}(1)$, require special attention for their manipulation in comparison to the other terms. Using the expression for $r_{x_1}(1)$, $r_{x_2}(1)$ etc. in (B4) one can show that

$$
r_x(1) = \frac{1}{N}\sum_{k=1}^{P}\sum_{j=1}^{P}\eta_k\eta_j p_k^1\Theta, \quad \Theta = \Theta_I + \Theta_{II}, \tag{B5}
$$

$$
\begin{aligned}
\Theta_I &= \frac{[1-p_k^Tp_j^T][(\lambda-1)-g(p_k)-g(p_j)+g(p_kp_j)]}{(1-p_kp_j)(1-p_k^T)(1-p_j^T)} \\
&\quad+\left[\frac{1-(p_kp_j)^{(R-1)}}{1-p_kp_j}\right]\left(\frac{1-p_k^{\lambda T}}{1-p_k^T}\right)\left(\frac{1-p_j^{\lambda T}}{1-p_j^T}\right) \\
\Theta_{II} &= \frac{p_j^T(p_kp_j)^{-l}}{1-p_j^T}[(\lambda-1)-g(p_j)]
\end{aligned}
$$

Note that in comparison with the expression for $r_x(0)$, an extra term Θ_{II} appears in $r_x(1)$. Similar to the derivation of $r_x(0)$ and $r_x(1)$, an expression for $r_x(2)$ can be obtained as

$$
r_x(2) = \frac{1}{N}\sum_{k=1}^{P}\sum_{j=1}^{P}\eta_k\eta_j p_k^2\Theta, \quad \Theta = \Theta_I + \Theta_{II}, \tag{B6}
$$

$$
\begin{aligned}
\Theta_I &= \frac{[1-p_k^Tp_j^T][(\lambda-1)-g(p_k)-g(p_j)+g(p_kp_j)]}{(1-p_kp_j)(1-p_k^T)(1-p_j^T)} \\
&\quad+\left[\frac{1-(p_kp_j)^{(R-2)}}{1-p_kp_j}\right]\left(\frac{1-p_k^{\lambda T}}{1-p_k^T}\right)\left(\frac{1-p_j^{\lambda T}}{1-p_j^T}\right) \\
\Theta_{II} &= \frac{p_j^T\left[(p_kp_j)^{-1}+(p_kp_j)^{-2}\right]}{1-p_j^T}[(\lambda-1)-g(p_j)]
\end{aligned}
$$

with $g(\cdot)$ being given in (B3). Finally, by observing the pattern that emerges in the expression for $r_x(0)$, $r_x(1)$, and $r_x(2)$, a general expression for $r_x(\tau)$ can be obtained as given by (2.17).