Enhancement-Techniques of Short-Duration Narrowband Signals with application to ABR Audiometry

Rajeev Agarwal

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy at

Concordia University

Montréal, Québec, Canada

October 1995

ISBN 0-612-05070-X

Canada

ABSTRACT

Enhancement Techniques of Short Duration Narrowband Signals with application to ABR Audiometry

Rajeev Agarwal

The detection estimation of rarely occurring transient signals is a frequently encountered problem in many areas. The problem can be broken down into two parts 1) Detection and 2) Estimation. This thesis addresses the second of the two parts. The underlying theme in this thesis is the estimation of transient signals embedded in noise, given their presence and *time-of-arrival*. In general we address the problem of estimating short-duration signals in the presence of noise.

Mathematical description or representation of signals ultimately outlines the methods employed in processing them. In this work, we consider two forms of signal modeling within the Structural Signal Processing framework. The relative advantages of the modern parametric (ARMA, AR, and MA) and the classical nonparametric series expansion model are exploited to generate null filters for processing such signals.

Within the parametric forms of null filtering, we consider the conventional Constrained Notch Filter. The deviations in frequency response of the CNFs caused by the pole-contraction factor ($\alpha < 1$) are shown to reduce the allpass nature of the filter. To quantify these errors, a new measure of *transparency misadjustment* is developed for the Almost Symmetrical-ARMA based filters. An expression of the *transparency misadjustment coefficient*, in terms of the filter coefficients, is derived. Further, it is shown that by increasing the order of the CNFs by strategic pole/zero pair placement, it is possible to correct for such distortions. Three closed-form methods for finding these pole/zero pair

location are derived. This improved performance is the allpass nature of the new filter can then be traded to obtain shorter transient durations for processing short duration signal

In the realm of nonparametric method we derive Statistically Optimal Null Filters (SONF) using a newly suggested Instantaneous Matched Filter as a key building block. To develop this approach we consider the maximum output signal to noise ratio and the least squares criteria. In this approach, we exploit the ability of the Matched Filter to process signals quickly, enabling the SONFs to process short duration signals. Several variants of the SONF are presented. Detailed analysis under certain constraints, indicating the equivalence of SONF's to Kalman Filter are also presented. Simulation results are used to verify/compare different SONFs and CNFs.

To further motivate the problem of processing short duration signals (transients) we consider the application of Evoked Responses to nonbehavioral hearing loss diagnosis of humans. The Auditory evoked Brainstem Response (ABRs) are elicited from the measured signals by averaging techniques. Often averaging does not adequately reveal the desired ABR. In the thesis, it is shown how averaged ABR waveforms can be processed using a combination of adaptive and recursive SONFs to minimize the residual noise. Several examples using real data are presented.

To my Late Father

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisors Drs. Eugene I. Plotkin and M. N. S. Swamy for their continual interest and guidance throughout this work. Their constructive criticism, advice and constant encouragement was immeasurable. This work is a culmination of their constant demand for high standards. It is with great pride that I acknowledge not one, but two great thesis supervisors.

I wish to express my sincere gratitude and heartfelt appreciation to Dr. H. J. Heki for introducing me to VBR Audiometry and the need for signal processing therein. At the initial stages of this work he was at the Department of Otolaryngology at the Royal Victoria Hospital, Montreal, Canada. Making his laboratory available for experiments was of great help.

It is with pleasure that I acknowledge the many discussions with friends and colleagues at the *Center for Signal Processing and Communications*. Special thanks go to Drs. Wen Long, Vladimir Shatalov and Y. Yoganandam for their interest and stimulating discussions.

I gratfully acknowledge the support of Natural Science and Engineering Research Council (NSERC) through a scholarship and Concordia University through Concordia University Fellowship. This work was also supported in part by NSERC and FCAR through grants awarded to Drs. E.I. Plotkin and M.N.S. Swamy.

I wish to thank my family for their constant support and encouragement. Special thanks go to my parents, without their confidence in me, this work would not be possible.

Last, but not the least, I would like to thank my wife, Nidhi, for her patience, constant encouragement and love through the course of this work and especially during the low points, it was a source of motivation to continue.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

ABR	Auditory evoked Brainstem Response
ACS	Autocorrelation Sequence
AR	Auto-regressive
ARMA	Autoregressive-Moving Average
ASTV-ARMA	Almost-Symmetrical Time-Varying ARMA
AWGN	Additive White Gaussian Noise
CANF	Constrained Adaptive Notch Filter
CNF	Constrained Notch Filter
CT	Computer Tomography
DFT	Discrete Fourier Transform
ECA	Error-Correction Approach
EEG	Electroencephlograph
ER	Evoked Response
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FM	Frequency Modulated
GS	Gram-Schmidt
KL	Karhunen-Loève
IIR	Infinite-Impulse-Response
IMF	Instantaneous Matched Filter
LMS	Least-Mean Square
LS	Least-Squares
LTI	Linear Time Invariant
LTV	Linear Time Varying
MA	Moving Average
MF	Matched Filter
MMSE	Minimum Mean-Square Error
MNF	Modified Notch Filter

MSE	Mean-Square Error
NB	Narrowband
NBW	Narrow Bandwidth
NC-SONF	Noncoherent Statistically Optimal Null Filter
NMR	Nuclear Magnetic Resonance
PSD	Power Spectral Density
SIR	Signal-to-Interference Ratio
SNR	Signal-to-noise Ratio
SNR_o	Output Signal-to-noise Ratio
SONF	Statistically Optimal Null Filter
SPZC	Symmetric Pole/Zero Constellation
SSP	Structural Signal Processing
SSR	Structural Signal Representation
TOA	Time-of-Arrival
α	Pole-Contraction or Symmetric factor
T_{ob}	Observation interval
$K_s(t,u)$	Covariance of signal $s(t)$
μ_i	ith Eigenvalue
$\phi_i(t)$	ith basis function
N_o	Power of white noise
β	damping factor
ω	Angular frequecny (rad/sec)
θ	Initial Phase
γ	Step-Size of adaptation
Γ	Transparency misadustment coefficient
diag(x)	Diagonal matrix composed from elements of x

Chapter 1

Introduction

The detection/estimation of rarely-occurring transient signal waveforms is a frequently encountered problem in many areas. Image processing for military and medical purposes, biological signal detection, radar and sonar detection, monitoring and diagnosis of machinery are but a few examples. Usually the existence or appearance of such waveforms is associated with an undesirable phenomenon, such as a failure in machinery. Their *time-of-arrival* (TOA) is unknown and to prevent damage (as in a military threat) earliest possible detection is of extreme importance. In literature there are many papers devoted to the estimation of TOA of transients, see for example [6]. In certain instances, we are not only interested in the detection, but also in the classification and estimation of these signals [7]. For example, in Evoked Response (ER) audiometry we know the TOA or the onset of the approximately known signal, it is required to first detect its presence and then to estimate the waveform for further diagnosis. The transient signal processing problem can, therefore, be separated into two parts:

- Detection of the presence of the transient signal and estimation of its TOA
- Estimation of the transient waveform, given its TOA.

A transient signal, in general, can be defined as a signal whose temporal duration is shorter than the observation interval. It is for this reason, very often the term "short or finite duration" and "transients" are used synonymously. Transients can be either deterministic or random. In the latter case, they are nonstationary [8]. In this thesis, we primarily deal with deterministic transient signals, although, our approaches are applicable to random signal as well. Transient signals are often nonsymmetric, possess abrupt changes and decay in amplitude with time — much like an exponential function. Exponential functions model the behavior typical to many physical transient phenomenon quite well. Additionally, such a model offers a representation of the transients as a sum of narrowband (NB) components.

The underlying theme of this thesis is the estimation of transient signals in the presence of noise given their presence and in some applications, their TOA. In general, we aim to tackle the problem of estimating short-duration signals in the presence of noise irrespective of whether they are of decaying type or not.

Certain parameters, such as the frequency characteristics of the transient signal, can be estimated from the received signal. Once the presence of signal is ascertained, various estimation methods of different degrees of sophistication can be used. These can be either linear or nonlinear. To the linear category we can attribute the short time Fourier Transform [9], the Gabor Transform [8] and the Wavelet Transform [10]. These fall into the category most commonly known as linear time-frequency analysis. Implicit in the FFT approaches is the assumption that the data is sufficiently well modeled as a sum of harmonically related sinusoids. In addition to its inherently poor frequency resolution (caused by the implicit data windowing) a broadening of spectral line will appear when FFT is applied to strongly damped sinusoids [11]. This results in the failure to resolve closely spaced damped sinusoidal components. To the nonlinear category, we can assign time frequency techniques such as smoothed periodogram [12], Wigner-Ville Transform [13, 14] and the ambiguity function [15]. These methods constitute nonparametric methods.

Processing of NB signals immersed in noise is a common problem in the signal processing literature to which the more modern parametric methods have been applied extensively. The problem of enhancement/suppression of such signals has been well addressed and falls in the category of adaptive line-enhancers (ALEs), the most popular of which has been the tapped-delay line structure, first proposed by Widrow [16] and is now termed the Finite-Impulse Response ALE (FIR-ALE). Since this seminal paper, much work has been done in the parametric model approach to solve this problem [17, 18, 19, 20, 21] and many different structures have been proposed [22, 23, 24]. The advantage of this scheme is its simplicity, however, large orders are required to achieve good performance.

In the last several years numerous researchers have investigated the NB signal problem by applying the Autoregressive-Moving Average (ARMA) model based notch filters. In [25] Friedlander proposed a generalized ARMA model for such problems. Rao and Kung [26] proposed a constrained ARMA filter specifically suited for the enhancement/suppression of sinusoids which they have called Constrained Adaptive Notch Filter (CANF). A modification of such a filter was suggested by Nehorai [27], where the number of coefficients have been reduced to n from $2n$ using mirror symmetric moving-average polynomial.

The problem of processing signal components with frequencies varying in time is of great importance in some applications. One possible method of dealing with such signals is to use the above CANFs [26, 27]. This, however, yields satisfactory results only when the frequency variation is slow. To adequately deal with these types of signals, an almost symmetrical time varying ARMA (ASTV-ARMA) model has been developed [28, 29, 30, 31]. Many variations of the ASTV-ARMA are presented in [32] that yield satisfactory performance.

Note that large order ALEs require long convergence time, and thus precludes their use for short-record length signals. As such, applications of conventional parametric methods to estimate transients and short-record length signals are very limited. We illustrate this point by considering an example. Let a finite-duration received signal consist of a sinusoid in noise, i.e. $x(t) = s(t) + n(t)$ where $s(t) = A\cos(\omega_o t + \theta_o)$ for $t \in [0, T]$. We wish to use a Constrained Notch Filter (CNF) [26] to suppress $s(t)$ (or equivalently to enhance $n(t)$). To achieve this objective with a reasonable degree of fidelity, it is necessary to choose the filter coefficients such that the filter passband response is as close to unity as possible. In so doing, the settling time or the filter transient duration becomes excessively long. If T, the duration of $s(t)$, is much shorter than the duration of the filter transient, it is clear that it is not possible to use the CNFs to separate $s(t)$ and $n(t)$.

In contrast, our research efforts are devoted to the processing of short record length signals for which steady-state approaches are not applicable. Generally speaking, estimation of short-duration signals is a difficult problem.

Mathematical description or representation of signals ultimately outline the methods employed in processing them. In this research investigation, we consider two forms of signal modeling within the Structural Signal Processing (SSP) framework [33]. The relative advantages of the more modern parametric model (ARMA, AR, and MA) and the classical nonparametric series expansion model are exploited to generate null filters for processing of short-duration signals.

The first part of the thesis deals with parametric forms of null filtering. The frequency response of a conventional CNF [26, 27] becomes ideal as the pole-contraction factor α approaches unity, while the transient duration approaches infinity. For practically useful values of α ($\alpha < 1$), the phase and magnitude distortions, so caused, can reduce the allpass

nature of CNFs. By increasing the order of the CNFs based on strategic pole/zero pair placement, it is possible to correct for such distortions. Several closed form methods for finding such strategic pole/zero pair locations are derived. This improved performance in the allpass nature of the filter can then be traded to obtain shorter transient durations for the processing of short-duration signals.

The second part of the thesis deals with statistically optimal nonparametric methods of null filtering using a newly suggested Instantaneous Matched Filter (IMF) as a key building block. It is interesting to note that conventional MFs are exclusively used for the detection purposes, most notably in communication applications. In spite of this, in our approach, we propose a modification of the MF and use it as a fundamental building block for the implementation of null filters. We combine the maximum output signal to noise ratio (SNR) criterion of the MF and the MSE criterion to generate the new approach that we call Statistically Optimal Null Filter (SONF). In the SONF approach we exploit the ability of the MF to process signals quickly and show that it can be used to process short duration signals. A key assumption in this approach is that the signal under consideration can be represented as a series based on *a priori* known composing functions. Due to the optimality (in the MSE sense) and time-varying nature of the SONF, it is anticipated that it is equivalent to a Kalman filter. This equivalence is discussed in detail in this part of the thesis. Hence, the SONFs can be interpreted as a new implementation of the Kalman Filter.

To further motivate the problem of enhancement/suppression of short duration signals (transients) embedded in noise, we consider a biomedical problem. The clinical application of ERs for nonbehavioral diagnosis of the hearing capacity of individuals is considered. Because it is a non-invasive technique and does not require subject cooperation, it is ideally suited for patients not easily testable by behavioral means, i.e. infants, comatose and the mentally handicapped.

ABRs are a form of electroencephlographic (EEG) recordings, difference being that they are acquired in a controlled manner. As the name suggests, ABRs are obtained by applying a known auditory stimulus to the ear and then measuring the transient response of the brain using surface electrodes pasted to the scalp. The response is a set of peaks and valleys in the 10 *ms* window at the onset of the stimulus application. Thus, if a subject does

not hear the auditory stimulus, the response will not conform to the standard pattern and the deduction is hearing loss for that stimulus intensity levels. In this manner, a clinician may acquire a set of waveforms at different stimulus intensity levels and determine the level of hearing loss (hearing threshold). There are other by product results from this type of testing, such as possible tumor detection [3].

It is important to understand the drawbacks or problems associated with this type of nonbehavioral testing. Surface electrodes not only pick-up the desired evoked responses, but also pick-up the background EEG. The power of this background EEG can be as high as 200 times greater than the desired ABR. Given the short observation interval of the ABRs, a major signal processing task is to extract or enhance the ABR (desired signal) in the presence of strong background EFG (noise).

Audiology literature is rich in the techniques used to extract the elicited response [34, 35, 36, 37]. However, none have proven to be satisfactory. Even with the existence of plethora of signal processing techniques, the most widely used technique is still the averaging of the waveform acquired by repeated stimulation at the same intensity level. If the number of responses in the averaging is reduced, as is done to reduce the length of time required to collect a set of ABRs, the noise is not sufficiently suppressed. A method to minimize this residual noise component in the short-duration averaged ABR is required. We apply the SONF approach to process the averaged ABRs to further enhance ABR signal in the presence of strong noisy environment.

Outline of the thesis

This thesis can be divided into three parts. In the first part, we consider the parametric model of CNFs and show how it can be modified to reduce the filter transient durations. Second part contains a new nonparametric SONF based on the maximum output SNR and the MSE criteria and in the third part we consider a biological signal processing problem using the SONFs. In all three parts the common theme is the processing of short-duration signals. The following is a chapter by chapter outline of this thesis.

Chapter 2 is used to discuss the importance of signal modeling and hence the approaches dictated by this modeling. Two forms of signal modeling in the context of

Structural Signal Processing (SSP) are presented namely, the classical nonparametric series expansion model and the more modern parametric ARMA, AR and the MA models. The goal here is not to give an exhaustive survey but to provide a foundation for the remaining chapters. In this chapter we also present generalized parametric null filters. To quantify the level of *transparency* of these null filters, a new measure of the transparency misadjustment is presented.

In Chapter 3, we show that the tradeoff between the settling time or the transient duration of the CNFs and the transparency characteristics (defined by the value of the pole contraction factor) may be handled by incorporating into the design a phase minimizing strategy. It is shown that by increasing the order of an n coefficient notch filter (by strategically adding pole/zero pairs), it is possible to improve the transparency property of the filter for a given transient duration or reduce the transient duration for a given transparency. Three methods of finding these pole/zero pairs are developed. Close form expressions for the location of these roots are presented. Improvements in the transparency and the transient durations are shown via simulation results.

In Chapter 4 we present the idea of combining the maximum output SNR and the MSE criteria to develop the new nonparametric null filter that we have termed Statistically Optimal Null Filter. Several variation of this filter are presented, i.e. suboptimal, locally optimal, globally optimal etc. Since the globally optimal SONF cannot be easily implemented, a discrete-time recursive SONF is developed. Analysis of the SONF shows that they may be considered, under certain conditions, as a Kalman Filter. It is shown that the SONF leads to an extremely simple implementation of the Kalman Filter.

In order to use the SONFs, it is essential that the basis functions in the expansion of the signal under consideration be known *a priori*. In the damped sinusoid case, the damping coefficients are generally not known *a priori*. For this reason we have also developed an adaptive SONF.

In Chapter 5 we present simulation results evaluating the different SONFs developed in Chapter 4. Comparisons of the parametric approaches of Chapter 3 and the SONFs are presented. Two examples of filtering time-varying signal with the SONF approach are provided to indicate its ability to process signals with time-varying parameters (i.e. FM signals). An application to separate highly correlated signals using a modified SONF is

discussed. Simulations show promising results. The performance of the adaptive SONE is also evaluated via simulations.

In Chapter 6, we present an application of the nonbehavioral hearing loss assessment using ABR Audiometry. It is shown how the averaged ABR waveforms can be processed by using a combination of the adaptive and recursive SONEs to minimize the residual noise. Several example results using real data are presented.

Finally, Chapter 7 contains some concluding remarks and directions for future work.

Chapter 2

Structural Signal Model: Generalized Almost-Symmetrical Structure

A human observer can be trained to be highly perceptive in extracting information from graphical representation of a signal. This method of signal processing is, however, apt to be qualitative and does not readily lend itself to quantitative analysis. An example of this is in the Auditory Evoked Brainstem Audiometry (ABR), where the tester is trained to interpret different aspects of the ABR waveform from which various diagnosis can be made. Different testers tend to interpret the waveform differently leading to a somewhat subjective or qualitative evaluation. More will be said on the ABR application in a later chapter. For quantitative or numerical analysis, under specific conditions that are of physical importance, it is critical that the signal be represented by some mathematical model. An important aspect of the signal is in the information that it carries and not in the average power or energy since most signals are assumed to represent some physical quantity [38]. Not all signals are amenable to mathematical representation and conversely, not all mathematical models are equally useful in disseminating information. It is possible that many models may represent the signal efficiently, however, some models may prove to be more useful than others in describing the information bearing aspect of the signal.

Another important reason for signal representation is the physical insight that it provides. This is an essential step for the signal processing engineers, since it simplifies the thought process and hence increases the generality with which the information can be extracted. There are a number of simple concepts such as orthogonality, signal dimensionality and time-bandwidth product that lead to useful indications of the signal capabilities. Signal processing engineer can therefore use these ideas for the initial judgment which can later form a foundation for exact analysis [39]. Here the emphasis is placed on the ability of the signal processing engineer to predict what is feasible and what is not, since without this a great deal of exact analysis would be a wasted on unprofitable approaches [39]. The importance of problem specific signal representation cannot be overemphasized and it is for this reason that we briefly explore signal representation possibilities.

In Section 2.1, we briefly describe the modeling of signals from a Structural Signal Processing (SSP) point of view. Mathematical description or representation of a signal ultimately outlines the methods employed in its processing. In the next two sections, we briefly discuss two forms of signal modeling in the Structured Signal Representation (SSR). The goal here is not to give an exhaustive survey but to briefly provide a foundation for

the remaining chapters. Section 2.2 describes the classical nonparametric series expansion model, while, the more modern parametric representations (ARMA, AR and MA) are presented in Section 2.3. Section 2.4, as precursor to the next chapter, generalizes the design of parametric null filters. Section 2.5 discusses a new measure of the allpass nature of the null filters,the *transparency misadjustment coefficient*.

2.1 Structural Signal Processing

The solution to any problem involving the processing of signals invariably consists of three key stages: 1) mathematical representation or model of the signal; 2) requirements on the result of processing; and 3) the selection of an appropriate algorithm. In this section, we are concerned with the first of these tasks, the representation of the signal [38].

The underlying requirement for any signal model is its ability to adequately represent the specific conditions of a class of problems that are to be analyzed. That is, the model must match the information content of the problem for which it is intended.

All signal models can be classified into two classes, the contour model and the structural model [38]. To the contour model group, we can assign any mathematical representation of the signal that can be expressed in the form of some function of a set of arguments. Power Spectral Density (PSD) is one such representation. To the second group, the structural model, we can assign models that combine two or more contour models of the same signal according to some given rule or law. Thus, a structural model can represent a functional relationship between different contour representations of the same signal.

In mathematical modeling of a signal, it is possible to follow any of the above two approaches. In this thesis, we will follow the linear path of the structural signal representation, the shaded path as outlined in Figure 2.1. It should be clarified that a structural model does not necessarily convey any more information than the contour model, but, it may shed a different light on the problem at hand, leading to new possibilities in constructing algorithms for the processing of signals. The field of Structural Signal Processing (SSP) forms an important sub-discipline in signal processing (SP). The notion of associating structural properties with signals may seem strange when compared to the physical

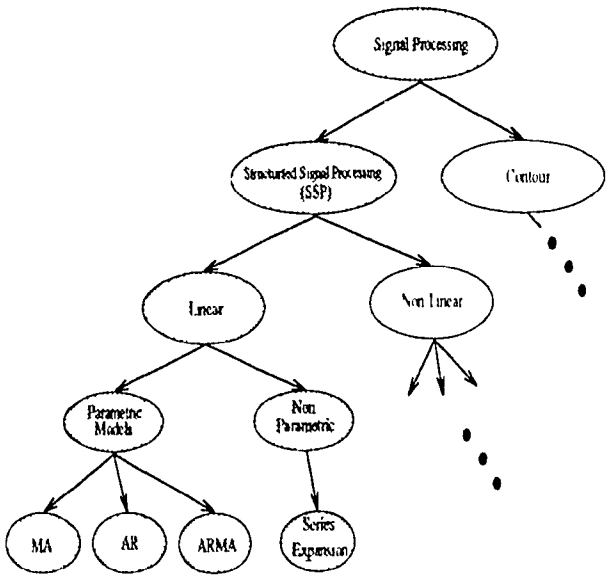


Figure 2.1: Structural Signal Processing framework.

properties of materials or buildings [40]. Here we are considering the characteristics or attributes that are not exactly defined or established for signals as they are for systems, i.e., linear, nonlinear, time-invariant or time-varying and so on. However, one does have some inclination as to what is meant by the term structure. The concept of SSP is based on the idea that the signal or process under analysis is a realization of an ensemble of several different transformations or conversions of the same signal. The characteristics of individual transformations may vary, however, there exist some intrinsic properties that belong to the ensemble of these transformations [33].

To illustrate this point, consider a signal $s(t, \underline{a})$, where $\underline{a}$ represents a vector of parameters and t is the independent variable (time). We can now perform specific transformations of $s(t, \underline{a})$ by the operators $\{L_k\}$ to obtain the set $\{s_k(t', \underline{b})\} = \{L_k s(t, \underline{a})\}$, where $\underline{b}$ is the new parameter vector and t' is the new independent variable. If this set of transformations is combined into an ensemble by an application of some specific rule or operator $\mathcal{G}$,

$$\mathcal{G}\left[\{s_k(t', \underline{b})\}\right] = \mathcal{G}\left[\{L_k s(t, \underline{a})\}\right] = f(\tau, \underline{\gamma}) \tag{2.1}$$

where $f(\tau, \underline{\gamma})$ is an *a priori* given function of the independent variable τ and $\underline{\gamma}$ is a vector of parameter, then this forms the basis of Structural Signal Representation (SSR). The

function $f(\tau, \underline{\gamma})$ does not depend on the signal $s(t, \underline{a})$ or any of its transformations. If $f(\tau, \underline{\gamma})$ is chosen to be 0 then the operator $\mathcal{G}$ is a Null Operator which characterizes the property of the ensemble $\{s_k(t', \underline{b})\}$. As an example consider the Nth order difference equation,

$$\sum_{0}^{N} a_k \Delta^{-k} s(t) = 0 \tag{2.2}$$

where Δ^{-k} is the kth order difference operator, the L_k operator in (2.1), and a_k's are constant weights. Each term $\Delta^{-k} s(t)$ represents a different transformation (realization) of $s(t)$, while the weighted sum of N such transformations represents the structural model in accordance with the definition of (2.1).

2.2 Non-Parametric Representation: Series Expansion

Classical form of signal representation is the series expansion of a signal. In contrast to the parametric model of the next section, the major properties of this type of modeling are related to the properties of the composing function. It is for this reason that we choose to classify them as nonparametric models.

The weighting coefficients of the basis functions, the a_k's in (2.2) depend on the signal and the chosen form of the composing functions. Some important issues are related to this kind of signal representation. Of special interest is the one where the chosen set of composing functions are orthogonal over an observation interval.

Let $s(t)$ be a finite energy (square integrable) signal defined over an interval $[0, T]$. It is well known that such a function can be expanded as a linear combination of some basis functions, $\phi_k(t), k = 0, 1, 2, \ldots N-1$, where in general N may approach infinity,

$$s(t) = \sum_{k=0}^{N-1} a_k \phi_k(t) \qquad \forall \quad 0 \le t \le T \tag{2.3}$$

The problem of selecting the appropriate basis function or the evaluation of the weighting coefficients has not been solved in general. However, numerous results exist for specific basis functions, which are used as an aid for making an appropriate choice in any given case.

One property that is desirable is the orthogonality of the basis functions. The

condition for orthogonality of basis functions can be stated as,

$$\int_0^T w(t)\phi_n(t)\phi_m^*(t) = \begin{cases} 0 & m \neq n \\ \mu_n & m = n \end{cases} \tag{2.4}$$

where $*$ represents complex conjugation, $w(t)$ is a weighting function and μ_n are real. If $\mu_n = 1$, then this set of basis functions is known as orthonormal.

As a result of this property, we can achieve what is known as the *finality of coefficients* which simply states that if the expansion of (2.3) is increased by one additional term, then the first N weighting coefficients remain unchanged. Another property arising from the orthogonality of the basis functions is related to the accuracy of the representation when less than the required number of basis functions are used. It states that by using basis functions that are orthogonal, we can achieve best approximation in the mean-square error sense. In general, to represent a signal exactly the number of basis functions required will be infinite while the corresponding values of the coefficients, a_k, become smaller and smaller as N increases. In practical situations, it is not possible to use infinite number of terms, hence, the resulting truncated expansion is inherently an approximation to the signal, $s(t)$,

$$\hat{s}(t) = \sum_{k=0}^{N-1} a_k \phi_k(t) \tag{2.5}$$

where $\hat{}$ indicates an approximation. The number of terms in (2.5) is determined by practical considerations. In the case where the minimum mean-square error

$$\overline{\epsilon^2} = \int_0^T [s(t) - \hat{s}(t)]^2 \, d\tau = 0 \tag{2.6}$$

the set $\{\phi_k(t)\}$ is known as *complete*.

Since orthogonal basis functions minimize the mean-square error and are convenient in terms of *finality of coefficients*, they are almost exclusively used in series expansions of signals.

Two of the most popular models for series expansions are the Fourier Series expansion for periodic signals based on a fixed set of complex sinusoids equally spaced in frequency, and the Shannon-Whitaker [41, 42] expansion based on a fixed set of sinc functions equally spaced in time.

At times, we have available a set of finite energy functions, $g_i(t)$, $i = 0, 1, \ldots M-1$ that may not be orthogonal over the interval $[0, T]$. In this case, by using the Gram Schmidt orthogonalization procedure [43], it is possible to orthogonalize these functions. The resulting number of orthogonal basis functions will be in general $N \leq M$.

Another important series expansion of a random process is known as the Karhunen Loève transform or expansion (KL expansion) [44]. In maintaining generality, we discuss the expansion of a random process that includes deterministic signals. The truncated version of the KL expansion has been shown to yield the best approximation of all expansions of the same dimension. As with other expansions it is computed by projecting the random process on a set of orthogonal basis. What makes this expansion interesting is that the basis functions need not be known *a priori* and are computed from the following integral equation

$$\int_0^T K_s(t, u)\phi_k(u)du = \mu_k\phi_k(t) \qquad \forall \quad 0 \leq t \leq T \tag{2.7}$$

where $K_s(t, u)$ is, in general, a time-varying covariance function of a process $s(t)$. The basis functions, $\phi_k(t), k = 0, 1, \ldots$, are known as the KL eigenfunctions and μ_k are the KL eigenvalues. The resulting signal, $s(t)$, can be written as

$$s(t) = \sum_{k=0}^{\infty} a_k\phi_k(t)$$

where a_k's are uncorrelated coefficients in the expansion. Two problems of practical considerations arise in employing this procedure to represent signals: 1) the covariance function may not be *a priori* known and 2) the solution to the eigen equation (2.7) is mathematically intractable in almost all but a few simple cases. In cases where the process $s(t)$ is periodic, the KL expansion is equivalent to the Fourier Series expansion.

2.3 Parametric Representation

The properties of the operators L_k and $\mathcal{G}$ determine whether SSR is linear, nonlinear, time varying or time-invariant, while the independent variable classifies it to be discrete or continuous [33]. Each method of SSR emphasizes different advantages or intrinsic properties of the signal being considered [33]. The simplest SSR model is one where the

operators L_k and $\mathcal{G}$ are chosen to be linear. For example,

$$\sum_{k=0}^{N} a_k L_k s(t) = 0 \tag{2.8}$$

where the weighted sum of the individual transforms $L_k s(t)$ (as in (2.1)) represents the operator $\mathcal{G}$ and a_k's are constant. This kind of a structural representation of $s(t)$ is referred to as the parametric model, and was first suggested by Baron de Prony in 1795 [15].

Current trends in processing of signals involve parametric approaches for both deterministic and random signals. More appropriately such approaches can be called model-based methods, since with these techniques a model has to be first *assumed or inferred* and then the parameters of the model are evaluated based on the observed signal. Figure 2.2 adequately describes the parametric approach. In the first block, the parameters of the chosen model are estimated. The second block uses these parameters and evaluates the estimate of the desired signal. The parameters effectively represent the signal. This is in stark contrast to the nonparametric approaches, where the parameters representing the signal are not directly evaluated. In this case, the signal of interest is estimated directly. The relationship between the power spectral density (PSD) and the autocorrelation sequence (ACS) can be considered as the nonparametric description of the second order statistics of a process. On the other hand, a parametric description may involve a time-series model of type (2.8), in which case the PSD is completely characterized by the parameters of the time-series, i.e, a_k's in (2.8) [12, 46].

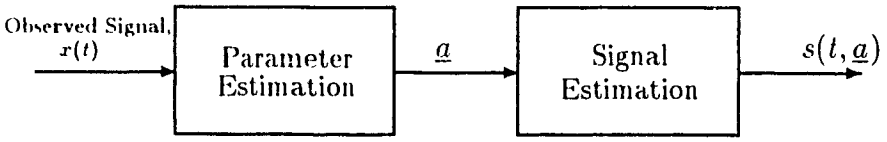


Figure 2.2: Procedure in Parametric Methods.

In the framework of SSP several forms of parametric signal representations are possible. A special class of Linear Structural Model includes the popular Autoregressive (AR), Moving-Average (MA) and the Autoregressive Moving-Average (ARMA) processes, where the models are represented by rational functions and are driven by white noise processes. With the advent of digital signal processing these models have found extensive use in

recent years. In the literature, the term parametric model has now become synonymous with these models exclusively. Hence, when discussing parametric models, we limit our discussion to such models.

If the building block, L_k in (2.8) is chosen to be the backward difference operator, i.e., $L_k s(n) = s(n-k) = s_k(n)$ with n as a discrete independent variable, then the model class is narrowed to the familiar difference equation,

$$\sum_{k=0}^{N_b} b_k x(n-k) + \sum_{k=0}^{N_a} a_k y(n-k) = 0 \tag{2.9}$$

or

$$y(n) = \sum_{k=0}^{N_b} b_k x(n-k) + \sum_{k=1}^{N_a} a_k y(n-k) \tag{2.10}$$

where $y(n)$ is the output sequence and $x(n)$ is the driving input and $a_0 = b_0 = 1$. It is evident that this difference equation can be considered as a particular structure model defined by (2.1) [33]. In the signal processing literature this model is called the ARMA model. Rewriting (2.10) using z-transforms,

$$Y(z) = \frac{B(z)}{A(z)} X(z) \tag{2.11}$$

where $A(z) = 1 + \sum_{k=1}^{N_a} a_k z^{-k}$ and $B(z) = 1 + \sum_{k=1}^{N_b} b_k z^{-k}$, we can specify the other two models, namely, the MA and the AR. If $A(z) = 1$, we obtain the MA model and with B(z) = 1 we have the AR model.

2.4 General Null Filters

The enhancement/suppression of NB signals is usually carried out by the Constrained Notch Filter, CNF. In this section, we briefly present the description of the design of the generalized Nth order null filters within the SSP framework. The discussion of CNFs is deferred to next chapter.

Two of the most important characteristics of null filters are the null singularities and transparency. These will be defined for general null filters. It is shown that to generate an ideal null filter that possesses each of these two characteristics, two different operators need to be combined in a cascade connection. In light of their function, these two operators have been termed the **null operator** for creating the null singularity and

its inverse the **anti-null operator** for maintaining transparency. Ideal null filters, so generated, are not practically implementable as the transient duration associated with them is excessively long. It is shown how the anti-null operator can be modified to make these filters implementable.

2.4.1 Null Operator

Hoene Wronski, in 1815, presented the fundamental results describing the necessary and sufficient conditions for a system of functions $\{g_1(t), g_2(t), \ldots, g_N(t)\}$ to be linearly independent solutions of a homogeneous differential equation (as presented in [47]). This condition has come to be known as the positiveness of the Nth order **Wronskian** of the set of functions $\{g_1(t), g_2(t), \ldots, g_N(t)\}$. Later, *Bortolotti* [48] extended these results to a generalized functional setting. He stated:

> *If $\{g_1(t), g_2(t), \ldots, g_N(t)\}$ be a system of analytic functions having a common domain of convergence, and* **L** *be a one-to-one mapping functional operation, then the necessary and sufficient condition for such a system of functions being associated with a homogeneous equation with respect to* **L** *is*

$$W_L^N(g_1(t), g_2(t), \ldots, g_N(t)) = \begin{vmatrix} g_1(t) & g_2(t) & \ldots & g_N(t) \\ Lg_1(t) & Lg_2(t) & \ldots & Lg_N(t) \\ \vdots & & & \\ L^N g_1(t) & L^N g_2(t) & \ldots & L^N g_N(t) \end{vmatrix} \neq 0 \quad (2.12)$$

This result provides a sound mathematical tool for linear structural processing, where the base operator **L** can be modeled as an arbitrary building block. If the operator **L** is chosen to be a differential operator, then W_L^N is known as the Wronskian. In this thesis, we will refer to (2.12) as the Wronskian *for any chosen operator* **L**.

It is well known that for N linearly independent functions $\{g_1(t), g_2(t) \ldots g_N(t)\}$, there exists a unique **monic homogeneous equation** of order N given by

$$\mathcal{G}[L_t[s(t)]] \equiv \frac{W_L^{N+1}(g_1(t), g_2(t) \ldots g_N(t), s(t))}{W_L^N(g_1(t) g_2(t) \ldots g_N(t))} = 0 \quad (2.13)$$

where W_L^{N+1} is an augmented Wronskian [17, 19] of the set $\{g_1(t), g_2(t) \ldots g_N(t)\}$.

$$W_L^{N+1}(g_1(t), g_2(t), \ldots, g_N(t), s(t)) = \begin{vmatrix} g_1(t) & g_2(t) & \ldots & g_N(t) & s(t) \\ Lg_1(t) & Lg_2(t) & \ldots & Lg_N(t) & Ls(t) \\ \vdots & & & & \\ L^N g_1(t) & L^N g_2(t) & \ldots & L^N g_N(t) & L^N s(t) \end{vmatrix} \neq 0 \tag{2.14}$$

Such a homogeneous equation also defines a linear operator $\mathcal{G}_I(s)$. Different building block operators, **L**, will result in different operators $\mathcal{G}_L(s)$. For example, Wronski [17] used differential building blocks while Casarati [18] used the difference operator. The operator $\mathcal{G}_L(s)$ of (2.13) can be explicitly written as a linear homogeneous equation

$$\left[1 + b_1(t)L + b_2(t)L^2 + \cdots + b_N(t)L^N\right] s(t) = 0 \tag{2.15}$$

where

$$b_i(t) = (-1)^{i+1} \frac{M_L^N\left(g_1(t)g_2(t)\ldots g_{i-1}(t), g_{i+1}(t)\ldots, g_N(t)\right)}{W_L^N(g_1(t)g_2(t)\ldots g_N(t))} \tag{2.16}$$

and M_L^N is the Minor of W_L^{N+1} with respect to the Laplace expansion of $s(t)$ [50]. Notice that we have used time-varying coefficients $b_i(t)$'s to maintain generality. If a signal $s(t)$ can be written as a linear combination of the complete set of functions $\{g_1(t), g_2(t) \ldots g_N(t)\}$, then (2.15) can be defined as an Nth order **null operator** and the set $\{g_1(t), g_2(t) \ldots g_N(t)\}$ constitutes an N-dimensional signal space, Ω. Such a signal space is the null singularity space of the null operator $\mathcal{G}_L(s)$.

2.4.2 Anti-Null Operator

Consider two signals, $s_1 \in \Omega$ and $s_2 \in \overline{\Omega}$ where $\Omega \cup \overline{\Omega} = I$ and $\Omega \cap \overline{\Omega} = \emptyset$. Our requirement is to define an operator or generalized null filter $\mathcal{Q}_L$ such that the following two objectives are satisfied,

$$\begin{aligned} \mathcal{Q}_L[s_1] &= 0 \\ \mathcal{Q}_L[s_2] &= s_2 \end{aligned} \tag{2.17}$$

for the signal enhancement/suppression problem. Of course, signal separation is one obvious application.

Given a set of functions $\{g_1(t)\ g_2(t)\ldots g_N(t)\}$ that span the subspace Ω, we can achieve the first of these two requirements using the $\mathcal{G}$ operator as described above. However, any signal s_2 that does not belong to the subspace Ω will be distorted by the operator i.e., $\mathcal{G}_L[s_1 + s_2] = \hat{s}_2 \neq s_2$. In order to achieve our second objective, it is necessary to find an operator $\mathcal{G}_L^\dagger$ that will reconstruct the signal s_2 without altering the null singularity of $\mathcal{G}_L$. By combining these two operators in a cascade connection, it is possible to construct an operator $\mathcal{Q}_L$ that meets both of the objectives in (2.17).

Definition 2.1 *The* transparency *of any linear operator $\mathcal{Q}_L$ (as defined above) implies that any signal not belonging to its null singularity subspace, Ω, passes through this operator undistorted. That is, if $s \in \overline{\Omega}$ and $\Omega \cap \overline{\Omega} = \emptyset$, then $\mathcal{Q}_L[s] = s$.*

Definition 2.2 *An anti-null operator $\mathcal{G}_L^\dagger$ is an inverse operator such that $\mathcal{Q}_L = \mathcal{G}_L\mathcal{G}_L^\dagger$ and $\mathcal{Q}_L$ maintains null singularity of $\mathcal{G}_L$ while satisfying Definition 2.1.*

Theorem 2.1 *The necessary and sufficient conditions for a time-varying operator $\mathcal{Q}_L$ to be transparent to any signal other than those spanned by its null singularity is that the null and the anti-null operators are mutually symmetrical.*

Proof: See [32].

The condition of perfect symmetry of the null and the anti-null operators is a very convenient result, since it allows us to easily form an anti-null operator for the construction of a null filter without any additional effort. If, in general, a time-varying null operator with a null singularity subspace Ω is defined as

$$\mathcal{G}_L(\cdot) = \left[1 + \sum_{i=1}^{N} b_i(t)L^i\right](\cdot) \tag{2.18}$$

then it is possible to describe the anti-null operator as

$$\mathcal{G}_L^\dagger = \left[\frac{1}{1 + \sum_{i=1}^{N} b_i(t)L^i}\right](\cdot). \tag{2.19}$$

For any signal $x(t)$ of order N, the input-output relationship can be written in the generalized ARMA canonical form,

$$\begin{aligned} y(t) &= \left[\frac{1 + \sum_{i=1}^{N} b_i(t)L^i}{1 + \sum_{i=1}^{N} b_i(t)L^i}\right] x(t) \\ &= \mathcal{Q}_L(\cdot)x(t) \end{aligned} \tag{2.20}$$

Generality of such null filters lies in the fact that the symmetry condition is valid for any chosen building block operator L. We may, therefore, choose L to be the difference operator yielding a time-varying discrete ARMA model. A detailed discussion of such filters and their properties can be found in [32].

In a linear time-invariant system, the stability condition is described by the location of the system function poles. An LTI system is absolutely stable *iff* its poles lie inside the unit disc. Extension of these results to the LTV case is unfortunately not valid. In [32] some new results on the stability of discrete LTV ARMA model have been presented. Here, we will restrict ourselves to the case where the coefficients $b_i(t)$'s in (2.20) are time invariant, leading to an LTI null filter, defined as

$$Q_L = \frac{1 + \sum_{i=1}^{N} b_i L^i}{1 + \sum_{i=1}^{N} b_i L^i} \tag{2.21}$$

Clearly, such a perfectly symmetrical operator cannot be implemented in practice. The identical roots of the MA and the AR polynomials force an infinite duration transient period [51]. To achieve implementability of this operator, it is necessary to break the symmetry between the homogeneous and the nonhomogeneous parts of the operator Q_L. This can be achieved by modifying the anti-null operator $Q_L^{\dagger}$ such that the operator Q_L is almost symmetric,

$$Q_L^{\alpha} = \frac{1 + \sum_{i=1}^{N} b_i L^i}{1 + \sum_{i=1}^{N} \alpha^i b_i L^i} = \frac{1 + \sum_{i=1}^{N} b_i L^i}{1 + \sum_{i=1}^{N} a_i L^i} \tag{2.22}$$

where $\alpha \in [0, 1)$ is called the symmetric factor.

2.5 Transparency

The degree to which the operator Q_L^{α} of (2.22) is transparent to signals not belonging to its null singularity space is determined by the symmetric factor, α. In this section, we present a quantitative description of the measure of transparency.

Consider $x(t)$, a mixture of two signals $s_1 \in \Omega$ and $s_2 \in \overline{\Omega}$ where $\Omega \cup \bar{\Omega} = I$ and $\Omega \cap \overline{\Omega} = \emptyset$. According to the definition of transparency in the last section, the numerator or the MA polynomial of Q_L^{α} must suppress s_1 while the complete operator Q_L^{α} must be *transparent* to the signal s_2. By *transparent* it is meant that s_2 will pass through this

operator *undisturbed* [31]. The quality of this transparency depends on the value of α. As $\alpha \to 1$, the respective coefficients of the MA and AR parts of $\mathcal{Q}_L^i$ approach equality and the total operator approaches an ideal zero-phase allpass filter. Thus, the transparency becomes almost ideal, i.e., the output $y'(t)$ of the operator in Figure 2.3 is close to $s_2(t)$. In practical situations, however, values of α very close to unity cannot be used due to the associated long transient durations. Typically, values of α from .8 to .99 are used, leading to a non-ideal filter.

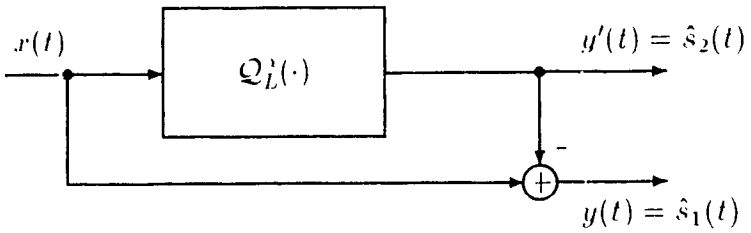


Figure 2.3: General null filter in a Line-Enhancement Configuration.

In what follows, we describe a new criterion for measuring the transparency imperfections and use this criterion to illustrate the improvements in the proposed filter characteristics. To do so, consider a realization of the ARMA that is similar to the direct form implementation, as in Figure 2.4, where we have introduced a second input $U_c(t)$. The operators $\mathcal{G}_L^{MA}(\cdot)$ and $\mathcal{G}_L^{AR}(\cdot)$ are respectively the MA and the AR parts of $\mathcal{Q}_L^a$. We call this artificial[1] signal, the complementary signal, which is additively applied to the input of the AR part to correct for transparency imperfections, i.e., to force the filter to maintain ideal transparency. The RMS value of this signal, normalized *w.r.t.* the input signal power, is taken as a measure of the transparency imperfection.

Theorem 2.2 *The measure of transparency of the filter defined by the operator $\mathcal{Q}_L^a$ of (2.22) for a zero-mean wide-sense-stationary input signal can be written as*

$$\Gamma = \sum_{i=1}^{N} K_i \tag{2.23}$$

where the K_i's are constants that depend on the linear operator L_i's and the filter coefficients $\{b_i, a_i, i = 1, 2, \ldots, N\}$.

[1] By *artificial* we mean that it is only a mathematical concept defined to describe the transparency imperfections, it does not specify a real (physical) signal.

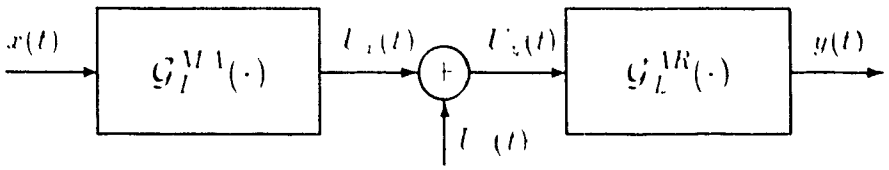


Figure 2.4: Direct form implementation of the general null filter

Proof: Let

$$G_L^{MA}(\cdot) = \left[1 + b_1 L + b_2 L^2 + \cdots + b_N L^N\right](\cdot) \tag{2.24}$$

and

$$G_L^{AR}(\cdot) = \left[1 + a_1 L + a_2 L^2 + \cdots + a_N L^N\right](\cdot) \tag{2.25}$$

represent the generalized MA and AR operators in in Figure 2.4 with $x(t)$ and $y(t)$ as input and output of the modified ARMA model, respectively. The input and output relationship of each of the two blocks in Figure 2.4 can be written as

$$U_x(t) = \sum_{i=0}^{N} b_i L^i x(t) \tag{2.26}$$

$$y(t) = \frac{1}{\sum_{i=0}^{N} a_i L^i} U_y(t) \tag{2.27}$$

where $a_o = b_o = 1$ and $U_y(t)$ and $U_x(t)$ are the input and output of the AR and MA parts, respectively. From Figure 2.4, the input of the AR part, $U_y(t)$, can also be written as

$$U_y(t) = U_x(t) + U_e(t). \tag{2.28}$$

Substituting (2.26) and (2.27) in (2.28) and solving for $U_e(t)$, we obtain

$$U_e(t) = \sum_{i=0}^{N} b_i L^i x(t) - a_i L^i y(t). \tag{2.29}$$

Ideal transparency implies that the input will propagate to the output undisturbed, i.e. $y(t) = x(t)$, in which case (2.29) becomes

$$U_e(t) = \sum_{i=0}^{N} (b_i - a_i) L^i x(t) = \sum_{i=1}^{N} c_i L^i x(t) \tag{2.30}$$

In equation (2.30), $U_c(t)$ is simply the output of the linear system defined as

$$\sum_{i=1}^{N} c_i L^i \overset{\mathcal{L}}{\longleftrightarrow} H(s) = \frac{P(s)}{Q(s)} \tag{2.31}$$

If the coefficients are assumed to be real, roots of $Q(s)$ are in the left half plane and $x(t)$ is Wide-Sense-Stationary (WSS) then $U_c(t)$ is also WSS. By using (2.31), the Laplace transform of $U_c(t)$ can be written as

$$\begin{aligned} S_U(s) &= S_x(s) H^+(s) H^-(s) \\ &= S_U^+(s) + S_U^-(s). \end{aligned} \tag{2.32}$$

where the superscripts '+' and '-' respectively represent the causal and anticausal parts of $H(s)$. Assuming that all the roots of $S_U^+(s)$ are simple and real, $S_U^+(s)$ may be expressed as

$$S_U^+(s) = \sum_{i=1}^{N} \frac{K_i}{s - s_i}. \tag{2.33}$$

The autocorrelation of $U_c(t)$ can now be written as

$$R_U^+(\tau) = \sum_{i=1}^{N} K_i e^{s_i \tau}. \tag{2.34}$$

Since $U_c(t)$ is a real process $R_U(\tau) = R_U(-\tau)$, the RMS value of $U_c(t)$ is

$$R_U(0) = \sum_{i=0}^{N} K_i \qquad \blacksquare \tag{2.35}$$

From (2.30) we note that the complementary signal is directly related to the filter coefficients. Normalizing the RMS value of the complementary signal w.r.t. the input signal yields the transparency coefficient

$$\Gamma = \frac{E\left[U_c^2(t)\right]}{E\left[U_x^2(t)\right]} \tag{2.36}$$

where $E[\cdot]$ is the statistical expectation for the case of a random input mixture. As the RMS value of $U_c(t)$ decreases, the filter becomes more and more transparent, that is, less and less energy is required to maintain the transparency of the filter. Assuming $x(t)$ is a zero-mean wide-sense-stationary process such that $E[L^i x(t) L^j x(t)] = 0$ for $i \neq j$ and $E[L^i x(t) L^j x(t)] = \sigma_x^2$ for $i = j$, where $i, j = 0, 1, 2, \ldots, N$, the transparency misadjustment coefficient can be written as a function of the filter coefficients in the form

$$\Gamma = \sum_{i=1}^{N} (b_i - a_i)^2. \tag{2.37}$$

From (2.22) and (2.37), it is clear that as $\alpha \to 1$, then $a_i \to b_i$ and $\Gamma = 0$, yielding a perfectly transparent filter. In short, transparency misadjustment is defined as the measure of imperfections in the allpass nature of the operator away from the singular points. At this point, we point out that the proposed measure of transparency imperfection may be extended to any filter of the almost symmetrical model type, including the time varying case.

Chapter 3

Time-Invariant Almost-Symmetrical Notch Filter and its Optimization

The problem of enhancement/suppression of narrowband (NB) signals in the presence of wideband interference has been well addressed in the past and the tool for this was the Adaptive Line-Enhancer (ALE), originally proposed by Widrow [16]. The basic blocks of the ALE are the tapped-delay line Finite-Impulse Response (FIR) filter and an adaptation rule such as the LMS criterion to adjust the filter weights. It is a multifaceted problem in that the estimate of the sinewave frequencies, the enhancement/suppression of the NB signals immersed in broadband (BB) signal or the tracking of signals may be desired.

In the last several decades numerous researchers have investigated the above problem by applying the autoregressive-moving average (ARMA) based notch filters, i.e. IIR filters in an effort to model the problem more exactly. Rao and Kung [26] proposed a constrained ARMA filter specifically designed for the suppression/enhancement of sinusoids. Their structure for an n-notch filter uses an ARMA $(2n,2n)$ model in which each pole/zero pair is constrained to lie on the same radial line such that the ith pole must lie between the origin and the ith zero on the unit circle. The closeness of the pole to the unit circle is defined using a pole-contraction factor α [26, 27]; as α approaches unity, the filter transparency[1] characteristic approaches the ideal one. A modification of such a filter was suggested by Nehorai [27] in which the number of filter coefficients were reduced to n from $2n$ using mirror-symmetric moving-average polynomial. An almost symmetric ARMA has also been considered in [52, 31] for the time-varying case. In literature these filters are generally referred to as Constrained Notch Filters.

In all the above designs of notch filters one considers only the magnitude response. That is, the design criterion used is unit amplitude at all frequencies other than the notch point, where it reaches is zero value. No constraint on the phase response has been incorporated into the design of such filters. Of course, to achieve the desired response, one can simply use a value of α close to unity. This, however, leads to a solution of no practical interest, since the transient duration becomes excessively long.

In this chapter we show that this tradeoff between the transient duration and the transparency characteristics (defined by the value of the pole-contraction factor) may be handled by incorporating into the design a phase-minimizing strategy. By increasing the order of an n-coefficient notch filter of [27] (by strategically adding pole/zero pairs), it is

[1]The quantitative measure of the transparency is introduced in the first section.

possible to force the phase response to rapidly approach zero away from the notch points. This significantly improves the transparency property of the filter for a given transient duration or reduces the transient duration for a given transparency.

The proposed method is based on addition of pole/zero pairs to the existing notch filter. In the context of the constrained notch filters of [26, 27], higher order implies a filter with multiple notches. *In the present context, however, increased order filter possesses a single notch.* The filter that we start with will be referred to as simply the notch filter (NF) and the resulting filter due to additional pole/zero pairs will be referred to as the *modified notch filter* (MNF).

The organization of this chapter is as follows. In the first section we describe the conventional CNF in terms of the general N-order null filter described in the previous chapter. We also give some definitions that will be used in subsequent sections. Section two describes how the idea of increasing the CNF order by strategically placing additional zeros can improve the performance of the CNF. In section three, we describe three methods of finding these zero locations. Section four contains some simulation results, while section five summarizes this chapter.

3.1 Constrained Notch Filters (CNFs)

An ideal notch filter is an allpass filter with ideal singularities and its frequency response can be written as

$$H(\omega) = \begin{cases} 0 & \omega = \omega_i \qquad i = 1, 2, \ldots, N \\ 1 & \omega \neq \omega_i \qquad i = 1, 2, \ldots, N \end{cases} \tag{3.1}$$

A $2N$th order CNF, on the other hand, can be constructed as a cascade of N second order sections

$$H(z) = \frac{\prod_{i=1}^{N}(1 + \theta_i z^{-1} + z^{-2})}{\prod_{i=1}^{N}(1 + \alpha\theta_i z^{-1} + \alpha^2 z^{-2})} \tag{3.2}$$

where $\theta_i = -2cos(\omega_i)$ is tuned to suppress a sinewave of frequency ω_i.

We can derive the CNF as a particular case of the earlier description of general Nth order null filters. Henceforth, we need only consider second order case as higher order filters can always be constructed from lower order ones. The choice of independent variable

determines whether the system of concern is analog or discrete. Here in accordance with the nature of the filter to be constructed, we choose the independent time variable to be discrete, i.e., $t = k$.

Consider a sinusoid $s(k) = A\cos(\omega k + \phi)$. Since this is a second order signal, it can be written in terms of two basis functions, namely the set $\{cos(\omega k), sin(\omega k)\}$. If we now choose the base operators, L^j in (2.22) to be the delay operators, Δ where $\Delta^{-m}x(k) = x(k-m)$, then the null operator $\mathcal{G}_\Delta(x)$, with $x(k)$ being the unknown input signal, can be written as a homogeneous equation ($q = \Delta$ in conformity to standard notation)

$$\mathcal{G}_\Delta(x) = \left[1 + b_1 q^{-1} + b_2 q^{-2}\right] x(k) \tag{3.3}$$

where in accordance with (2.16)

$$b_1 = \frac{\begin{vmatrix} \cos\omega k & \sin\omega k \\ \cos\omega(k-2) & \sin\omega(k-2) \end{vmatrix}}{\begin{vmatrix} \cos\omega k & \sin\omega k \\ \cos\omega(k-1) & \sin\omega(k-1) \end{vmatrix}} = -2\cos\omega \tag{3.4}$$

and

$$b_2 = \frac{\begin{vmatrix} \cos\omega k & \sin\omega k \\ \cos\omega(k-1) & \sin\omega(k-1) \end{vmatrix}}{\begin{vmatrix} \cos\omega k & \sin\omega k \\ \cos\omega(k-1) & \sin\omega(k-1) \end{vmatrix}} = 1 \tag{3.5}$$

Without any additional effort, the anti-null operator can be simply written as

$$\mathcal{G}^\dagger_\Delta(\cdot) = \frac{1}{1 + \alpha b_1 q^{-1} + \alpha^2 b_2 q^{-2}}(\cdot) \tag{3.6}$$

Combining the null and the anti-null operators in an input-output relationship (as was done in (2.20)), we have

$$y(k) = \mathcal{Q}^\alpha_\Delta(x(k)) = \frac{1 + b_1 q^{-1} + b_2 q^{-2}}{1 + \alpha b_1 q^{-1} + \alpha^2 b_2 q^{-2}} x(k) = \frac{1 + b_1 q^{-1} + b_2 q^{-2}}{1 + a_1 q^{-1} + a_2 q^{-2}} x(k) \tag{3.7}$$

which is exactly the second order CNF of [26, 27].

3.1.1 Transparency of Constrained Notch Filters

One of the simplest and most commonly used forms of measure of deviation of the frequency response $H(\omega)$ from that of the ideal notch filter is the mean-squared error given by

$$\bar{\epsilon}^2 = \|1 - H(\omega)\|_2^2 = \frac{1}{\pi}\int_0^\pi \left[1 - 2 \mid H(\omega) \mid \cos\phi(\omega) + \mid H(\omega) \mid^2\right] d\omega \tag{3.8}$$

where $\|\cdot\|_2$ refers to the norm in L_2. $\overline{\epsilon^2}$ is easily seen to be minimized and reaches zero if the phase response $\phi(\omega)$ is zero and the magnitude is unity (except at the notch points) at all frequencies.

Although the above $\bar{\epsilon}^2$ is a good indicator of the total error (*transparency* imperfections) in the line-enhancement configuration of $H(\omega)$, it fails to describe the deviations of the frequency response of the nonideal filter away from the singular points (notch points of $H(\omega)$). The dynamic range of the magnitude response of $(1 - H)$ varies from around unity near the singular points to values lower than 10^{-6} increasingly away from these points. Thus, by averaging (L_2-norm), the behavior of the filter away from the singular points is masked. It is therefore necessary to find an alternative performance criterion that is able to characterize the intrinsic properties of the filter at the notch frequency as well as at any other frequency.

In what follows, we evaluate the *transparency misadjustment coefficient* of the last section for the particular case of CNFs. We can evaluate the expression for Γ, the *transparency misadjustment coefficient*, directly from (2.30) where the operator L is the delay operator,

$$U_c(t) = \sum_{i=0}^{N}(b_i - a_i)q^{-i}x(t). \tag{3.9}$$

From (3.9) we note that the complementary signal is directly related to the filter coefficients and the operator q^{-1}. Normalizing the RMS value of $U_c(t)$ w.r.t. the input signal power yields the transparency misadjustment coefficient

$$\Gamma = \frac{E\left[U_c^2(t)\right]}{E\left[U_x^2(t)\right]} \tag{3.10}$$

where $E[\cdot]$ is the statistical expectation for the case of a random input mixture. Assuming $x(t)$ is a zero-mean wide-sense-stationary process such that $E[x(t-i)x(t-j)] = 0$ for

$i \neq j$ and $E[x(t-i)x(t-j)] = \sigma_x^2$ for $i = j$, where $i, j = 0, 1, 2, \ldots, N$, the transparency misadjustment coefficient can be written as a function of the filter coefficient in the form

$$\Gamma = \sum_{i=1}^{N} (b_i - a_i)^2 \tag{3.11}$$

From (3.7) and (3.11), it is clear that as $\alpha \to 1$, we have $a_i \to b_i$ and $\Gamma \to 0$, yielding a perfectly transparent filter.

In addition to the above, we will also consider the following characteristics of the notch filter to evaluate its performance.

Definition 3.1 The notch bandwidth (NBW) is defined as $\Delta\omega = \omega_{3dB}^{+} - \omega_{3dB}^{-}$, where ω_{3dB}^{+} and ω_{3dB}^{-} are the upper and lower half-power point frequencies (i.e., $-3dB$ frequencies), respectively.

Definition 3.2 The stopband error (ϵ_{STOP}) is defined as

$$\epsilon_{STOP} = \int_{\omega_{3dB}^{-}}^{\omega_{3dB}^{+}} | 1 - H(\omega) |^2 d\omega \tag{3.12}$$

Definition 3.3 The passband error (ϵ_{PASS}) is defined as

$$\epsilon_{PASS} = \int_{0}^{\omega_{3dB}^{-}} | 1 - H(\omega) |^2 d\omega + \int_{\omega_{3dB}^{+}}^{\pi} | 1 - H(\omega) |^2 d\omega \tag{3.13}$$

3.1.2 Transient

Often we wish to process a signal of finite duration. Thus, the issue of transient duration is of critical importance; we further motivate the design of MNFs by taking into consideration the processing of transient signals. Consider a standard CNF of [26], $H(z, \alpha)$, with a given pole-contraction factor $\alpha = \alpha_o$. Since the pole-contraction factor α controls the location of the poles, it also controls the transient duration of the filter [51]. In the next section, we show that it is possible to design a higher-order MNF $H'(z, \alpha')$ with a pole-contraction factor $\alpha' < \alpha_o$ and a transient duration $\tau' = \tau_o$ such that the resulting imperfections in the characteristics becomes significantly lower. These improvements can now be traded to obtain a shorter transient duration by further decreasing the value of α' (i.e. increasing the pole/zero distance) to the point where the errors in the proposed MNF and the standard CNF are approximately equal. As a result, a reduction in the transient duration τ' can

be observed such that it becomes significantly lower than τ_o. Hence MNFs can also be viewed as an improvement in the transient behavior of constrained notch filters. Clearly, such filters with shorter transient durations can be useful in processing signals of finite durations.

It is clear that as α increases, τ also increases [26, 27]. Thus, in comparing MNFs against standard second-order notch filters, the value of α must be adjusted such that both filters have the same transient durations for fair comparisons to be made.

Definition 3.4 The transient duration τ is defined as the time taken for the step response of any notch filter to reach within $\pm 0.01\%$ of its steady-state response.

3.2 Modified Notch Filter

A necessary property of the notch filter is to have the zeros lie on the unit circle. This is achieved by the mirror symmetry of the numerator polynomial coefficients of the ARMA model of (3.2). A further constraint of the notch filter is to maintain the allpass nature at all frequencies other than the notch points defined by the zeros on the unit circle. This can be achieved by forcing the ith pole to lie on the same radial line as the ith zero and close to but inside the unit circle. For values of $\alpha \to 1$, the IIR filter of (3.2) is such a structure and may be realized as a cascade of second order sections, where each section is used to suppress a single sinusoid. Since by using first and second order filters one can represent a notch filter with any number of notches, in what follows, we need only consider the first and second order filters for improving the phase and magnitude response of the notch filter.

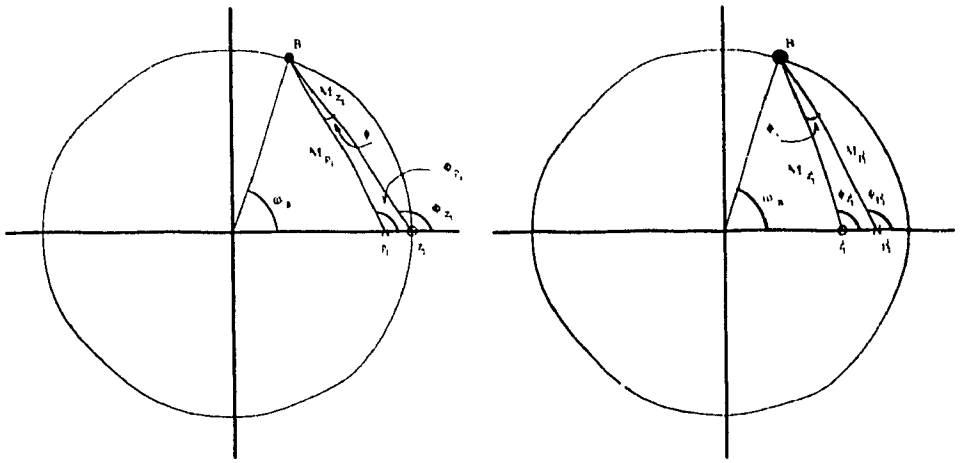
The first order notch filter which can be used for the suppression of any DC signal is represented by

$$H(z) = \frac{1 + z_1 z^{-1}}{1 + p_1 z^{-1}} \tag{3.14}$$

where $z_1 = -1$ and $p_1 = \alpha z_1$. The frequency response can be written as

$$H(e^{j\omega}) = \frac{1 + z_1 \cos\omega - j z_1 \sin\omega}{1 + p_1 \cos\omega - j p_1 \sin\omega} = \frac{M_{z_1}(\omega) e^{j\phi_{z_1}(\omega)}}{M_{p_1}(\omega) e^{j\phi_{p_1}(\omega)}} = M(\omega) e^{j\phi(\omega)}. \tag{3.15}$$

Consider the phasor diagram of Fig 3a for an arbitrary point B on the unit circle; we can determine the phase, $\phi(\omega_B)$, and magnitude, $M(\omega_B)$, responses at $\omega = \omega_B$ as $\phi = \phi_{z_1} - \phi_{p_1}$



(a) (b)

Figure 3.1: Phasor diagram of magnitude of response at $\omega = \omega_B$

and $M = M_{z_1}/M_{p_1}$, respectively. From the geometry in Figure 3.1a, we find that $\phi > 0$ and $M > 1$. *It is clear that the filter does not have a unity gain nor a zero phase response.*

The first order filter (single pole/zero pair) of (3.14) gives a positive phase and a magnitude greater than unity. Now let us extend the order of (3.14) by adding a pole/zero pair such that desired response is achieved. For filters with real coefficients, the poles and zeros must occur in complex-conjugate pairs or be real. Hence, if a single pole/zero pair is added to (3.14), then it must be constrained to be on the real line. If we now increase the order of (3.14) by an additional compensation pole/zero pair as in the phasor diagram of Fig 3.1b where the phase and amplitude response are $\phi_{z_1'} - \phi_{p_1'} < 0$ $M_{z_1'}/M_{p_1'} < 1$, respectively, *then it is possible to obtain a filter that has a resultant zero phase and a unit amplitude response*, satisfying

$$\phi + \phi_c = 0 \quad \textit{and} \quad \frac{M_{z_1}}{M_{p_1}} \frac{M_{z_1'}}{M_{p_1'}} = 1. \tag{3.16}$$

Note that the MNF retains the notch point of the original filter due to the zero on the unit circle, while improving the allpass nature away from the notch point due to the presence of compensation pole/zero pair. It is only required to strategically place the additional pole/zero pair. It is easy to see that this idea can be extended to any second order notch filter.

A second order notch filter can be represented as a cascade of two first order filters

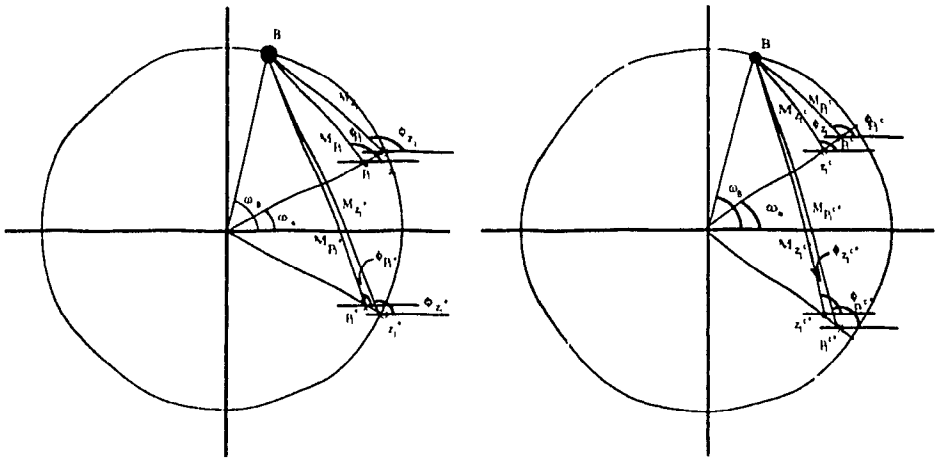


Figure 3.2: Phasor diagram of magnitude of response at $\omega = \omega_B$

in the form

$$\begin{aligned} H(z) &= \frac{1 + \theta z^{-1} + z^{-2}}{1 + \alpha\theta z^{-1} + \alpha^2 z^{-2}} \\ &= \left(\frac{1 + \gamma z^{-1}}{1 + \alpha\gamma z^{-1}}\right)\left(\frac{1 + \gamma^* z^{-1}}{1 + \alpha\gamma^* z^{-1}}\right) = \left(\frac{1 + z_1 z^{-1}}{1 + p_1 z^{-1}}\right)\left(\frac{1 + z_1^* z^{-1}}{1 + p_1^* z^{-1}}\right) \end{aligned} \tag{3.17}$$

where γ is a complex number and * denotes complex-conjugation. Figure 3.2a shows its phasor diagram at an arbitrary point B. If each of the two pole/zero pair is as in (3.17), then it is possible to use the earlier first order filter description of the phase and magnitude compensation. Mathematically, each part of (3.17) can be considered to be a generalization of (3.14), where instead of DC, each section is used to suppress a complex exponential. Figure 3.2b shows a phasor diagram at $\omega = \omega_B$ for an arbitrary compensation pole/zero pair location and their complex-conjugates. The compensation pole/zero pair, z_1^c and p_1^c, need be found such that the resultant phase and magnitude response follow as in the DC case. If this suitable compensation pole/zero pair can be found for one of the two sections of (3.17), then for the other section, its complex conjugate pair must be used to ensure that the MNF has real coefficients.

The filters generated from the two pole/zero patterns of Figure 3.1a and 3.1b are cascaded to form the new MNF that is used to suppress the DC signal

$$H'(z) = \left(\frac{1 + z_1 z^{-1}}{1 + p_1 z^{-1}}\right)\left(\frac{1 + z_1^c z^{-1}}{1 + p_1^c z^{-1}}\right) \tag{3.18}$$

or in the frequency domain

$$H'(\omega) = \left(\frac{M_{z_1}(\omega)}{M_{p_1}(\omega)} e^{j\phi(\omega)}\right)\left(\frac{M_{z_1^c}(\omega)}{M_{p_1^c}(\omega)} e^{j\phi_c(\omega)}\right) \tag{3.19}$$

The desired resultant phase and magnitude response at an arbitrary point B in Figure 3.1 are defined by (3.16). It is clear that if we choose z_1^c and p_1^c such that $\phi_c = -\phi$, $M_{z_1} = M_{p_1^c}$ and $M_{p_1} = M_{z_1^c}$, then the above two equations are satisfied. If these same z_1^c and p_1^c were to satisfy (3.16) for all points on the unit circle of Figure 3.1 other than $\omega = 0$, then this leads to an ideal notch filter. For a single compensation pole/zero pair in (3.18), it is only possible to achieve a finite improvement in the frequency response and as the number of compensation pole/zero pairs increase, the new filter $H'(z)$ begins to approach the ideal one.

We proceed by further constraining the poles of the compensation pole/zero pairs to be the same as the original filter pole, i.e. $p_1^c = p_1$ for $i = 1, 2, \ldots, N-1$. Such a constraint serves two purposes; first, we obtain a stable MNF (all poles are located inside the unit circle) and secondly, it is now required to only evaluate the location of the zeros of the increased order filter which significantly reduces the design complexity.

In the case where the compensation poles are set to p_1 by the above constraint, the increased Nth order DC MNF of (3.18) can be expressed as

$$H'(z) = \frac{(1+z_1 z^{-1})(1+z_1^c z^{-1})\cdots(1+z_{N-1}^c z^{-1})}{(1+p_1 z^{-1})^N}. \tag{3.20}$$

It is now only necessary to evaluate the values of the $N-1$ compensation zeros, $z_1^c, z_2^c, \ldots, z_{N-1}^c$ such that the resulting filter response becomes close to the desired response.

The above argument is equally valid for a second order filter (a filter with a notch at $\omega = \omega_o$), since it is composed of two first order sections. Thus in the same manner, the second order MNF can be represented in the form

$$H'(z) = \frac{(1+z_1 z^{-1})(1+z_1^* z^{-1})(1+z_1^c z^{-1})(1+z_1^{c*} z^{-1})\cdots(1+z_{N-1}^c z^{-1})(1+z_{N-1}^{c*} z^{-1})}{(1+p_1 z^{-1})^N (1+p_1^* z^{-1})^N} \tag{3.21}$$

where N is an even number. This latter requirement is necessary due to the complex conjugate pair necessity which ensures real filter coefficients. As N increases, a better approximation to the ideal filter response can be achieved.

3.3 Evaluation of the compensation roots

The 2Nth order Modified Notch Filter, MNF, can be written as a cascade of the CNF and a compensation section

$$H'(z) = H(z)H_c(z) = \frac{B(z)}{A(z)} \cdot \frac{\tilde{B}_c(z)}{\tilde{A}_c(z)} \tag{3.22}$$

where the prime indicates the MNF, $H(z) = B(z)/A(z)$ is the second order CNF and $H_c(z) = \tilde{B}_c(z)/\tilde{A}_c(z)$ is the 2N-2 order compensation section. In [53, 54], the poles of the compensation section are constrained to be the same as the original poles of the CNF. That is, if the complex conjugate pair (z_p *and* z_p^*) are the poles of the CNF then (z_p^i *and* z_p^{i*}) are the poles of $H_c(z)$ (the roots of $\tilde{A}_c(z)$), where $z_p^i = z_p$ for $i = 1, 2, \ldots N-1$.

In this section, we describe methods of evaluating the location of the compensation zeros needed for the design of MNFs. In describing these techniques, we will restrict the problem to a second order filter (a notch filter for the suppression of a single sinusoid).

3.3.1 Error-Correction Approach (ECA)

Consider a second order constrained notch filter given by (3.17). Note that by simply cascading N such identical sections, the overall amplitude response of the resulting notch filter will actually worsen as the number of sections in the cascade increases. In fact, since the original constraint of [26, 27] on the pole location causes deviations in the frequency response, increase in the distortion of the cascaded structure is only too obvious.

In view of the compensation strategy described earlier, we wish to increase the order of the filter such that the addition of the zeros provides phase and magnitude compensation. In this method, we first design a higher order line enhancer (LE) $H_{LE}(\omega)$ from which we derive the transfer function of the MNF as, $H'(\omega) = 1 - H_{LE}(\omega)$. Frequency response of an ideal-line enhancer is

$$H_{LE}^{IDEAL}(\omega) = \begin{cases} 1 & \omega = \omega_i \qquad i = 1, 2, \ldots, N \\ 0 & \omega \neq \omega_i \qquad i = 1, 2, \ldots, N \end{cases} \tag{3.23}$$

and is shown in Figure 3.3.

If an LE can be designed such that the frequency response $H_{LE}(\omega)$ is close to the

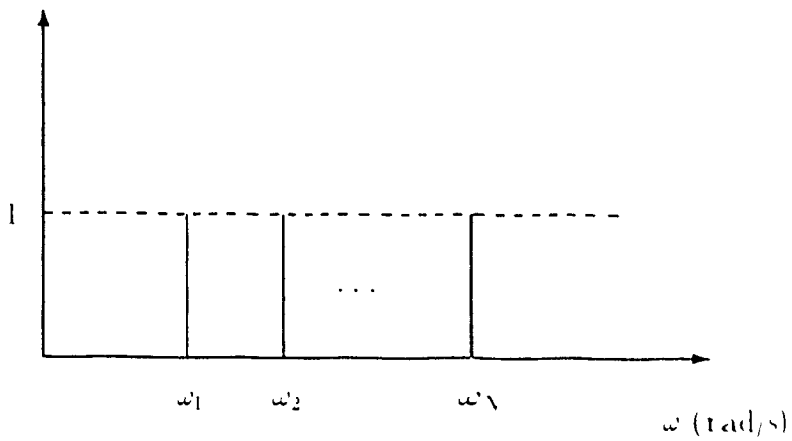


Figure 3.3: Ideal line enhancer of Nth order

ideal one, then the design of the MNF is completed. The transfer function of a line enhancer corresponding to $H(z)$ of (3.17) is

$$H_{LE}(z) = 1 - H(z) = \frac{(\alpha - 1)\theta z^{-1} + (\alpha^2 - 1)z^{-2}}{1 + \alpha\theta z^{-1} + \alpha^2 z^{-2}} \tag{3.24}$$

Frequency responses of a cascade of one, two and four such filters are shown in Figure 3.4, (i.e, $N = 1, 2$ and 4). Figure 3.4 indicates an improvement in the LE behavior as the number of sections in the cascade is increased. The MNF can now be formulated as

$$H'(z) = 1 - [H_{LE}(z)]^N \tag{3.25}$$

where $2N$ is the resulting MNF order. As N increases, the response gets closer and closer to the ideal one. It is important to realize that the cascade of LEs in (3.24) preserve a unit amplitude and zero phase at the notch frequency defined by θ. This condition can easily be verified. We illustrate this method with an example by redesigning a second order notch filter to form a fourth order MNF. To distinguish this MNF from that of the direct pole/zero placement method (described next), we use the subscripts 'ECA' in $H'_{ECA}(z)$.

Example 3.1 Line-enhancer associated with the second order notch filter is represented by (3.24). From this $H'_{ECA}(z)$ is formed to be (by choosing N to be 2 in (3.25))

$$\begin{aligned} H'_{ECA}(z) &= 1 - [H_{LE}(z)]^2 \\ &= \frac{1 + 2\theta\alpha z^{-1} + (2\alpha^2 + 2\alpha\theta^2 - \theta^2)z^{-2} + 2\theta(\alpha^2 + \alpha - 1)z^{-3} + (2\alpha^2 - 1)z^{-4}}{1 + 2\theta\alpha z^{-1} + \alpha^2(2 + \theta^2)z^{-2} + 2\theta\alpha^3 + \alpha^4 z^{-4}} \end{aligned} \tag{3.26}$$

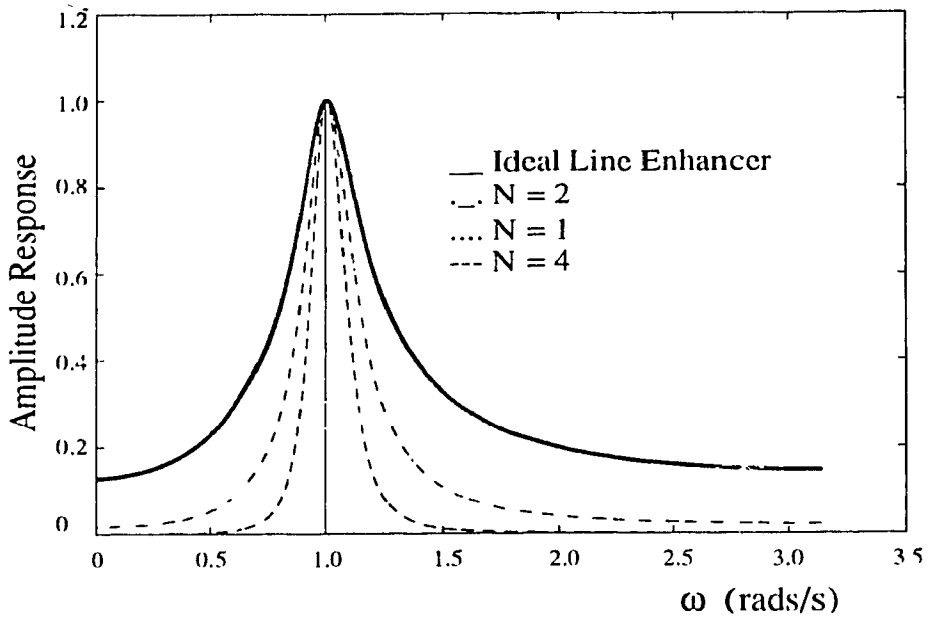


Figure 3.4: Line-enhancer magnitude response.

The notch point is defined by the value $\theta = -2\cos\omega_o$, where ω_o is the normalized notch frequency in the range $[0, \pi]$. Figure 3.5a shows the pole/zero pattern of $H'_{ECA}(z)$, while Figures 3.5b and 3.5c show the pole/zero patterns for the sixth and eighth order MNF, respectively.

3.3.2 Symmetric Pole/Zero Constellation (SPZC)

As in the previous design, the compensation poles are constrained to be at the same location as the original second order filter poles. We need only design a method to find the location of the compensation zeros. Examining the zero locations corresponding to the ECA method in Figure 3.5, it is clearly seen that as the order of the proposed $H'_{ECA}(z)$ increases, the zeros in the three pole/zero configurations are clustered very closely around an inner-circle of radius $(1-\alpha)$ with the centers located at the pole locations. It is evident, that in each pole/zero cluster of Figure 3.5, one and only one zero belongs to both the inner-circle and the unit circle. It is this zero that is necessary to maintain a notch at $\omega = \omega_o$. *The remainder of the zeros may or may not be exactly on this inner-circle, however, they are very close to it.* Furthermore, there appears to be some form of symmetry between the zeros in each cluster. We use this symmetry to design a closed form MNF based on

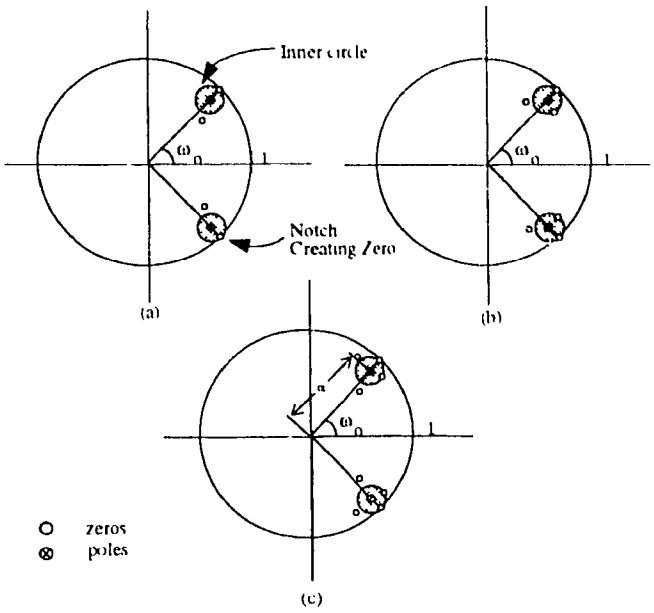


Figure 3.5: Pole/zero plot of ECA based MNFs. (a) fourth order; (b) sixth order; (c) eighth order.

the pole/zero placement. MNFs can be *designed by forcing the compensation zeros to be equally spaced on the inner-circle relative to the notch point zero* (i.e. $2\pi/N$ radians apart on the inner-circle, starting at the notch creating zero), creating symmetry much like the ECA method. Half of the zero locations of the MNF are given by:

$$z_i = (1-\alpha)e^{j\omega_o + i4\pi/N)} + \alpha e^{j\omega_o}, \qquad i = 0, 1, \ldots, N-1 \tag{3.27}$$

where ω_o is the notch frequency and the other half of the zeros are chosen to be complex conjugates of (3.27) to ensure a real coefficient filter. Equation (3.27) also ensures that for $i = 0$, z_o creates a notch at $\omega = \omega_o$ as required. This zero corresponds to the original second order part of the MNF and need not be determined here. Figure 3.6 shows the resulting pole/zero constellation for a sixth order MNF so designed.

The SPZC-based MNF can now be written as

$$H'_{SPZC}(z) = \frac{(1+z_o z^{-1})(1+z_o^* z z^{-1})(1+z_1 z^{-1})(1+z_1^* z^{-1})\cdots(1+z_N z^{-1})(1+z_N^* z^{-1})}{(1+\alpha z_o z^{-1})^N (1+\alpha z_o^* z^{-1})^N}. \tag{3.28}$$

The coefficients of the SPZC-based MNF can be obtained by expanding the numerator and the denominator of (3.28). The factorized polynomials in the numerator of (3.28) can be expanded using the following recursive relationship.

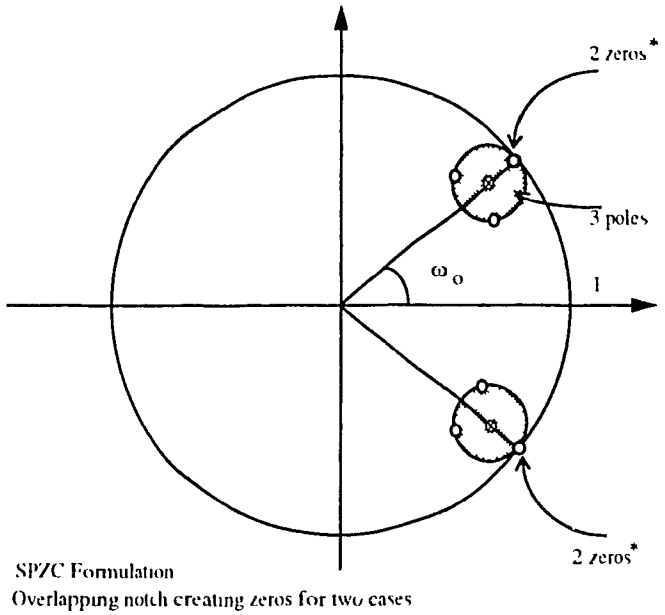


Figure 3.6: Pole/zero constellation for a sixth order SPZC-MNF.

It is always possible to write the factorized polynomial of degree $2N$ in the numerator of (3.28) as a product of N polynomials of degree two as

$$P(z) = (1 + \gamma_o z^{-1} + \beta_o z^{-2})(1 + \gamma_1 z^{-1} + \beta_1 z^{-2}) \cdots (1 + \gamma_{N-1} z^{-1} + \beta_{N-1} z^{-2}) \qquad (3.29)$$

where $\gamma_k = (z_k + z_k^*)$ and $\beta_k = \mid z_k \mid^2$ for $k = 0, 1, \ldots, N-1$. Equation (3.29) can now be expanded to generate a polynomial of degree $2N$. Let a_i^{2N-2} be the coefficients in the expansion of the first $N-1$ second order sections in (3.29) where $i = 1, 2, \ldots, N-1$ and $a_o^{2N-2} = 1$. Then the following recursive relationship can be used to derive the coefficients of the polynomial resulting from an expansion of N such second order sections:

$$a_i^{2N} = a_i^{2N-2} + \gamma_{N-1} a_{i-1}^{2N-2} + \beta_{N-1}^{2N-2}, \quad i = 1, 2, \ldots, 2N-2 \qquad (3.30)$$

$$a_{2N-1}^{2N} = \gamma_{N-1} a_{2N-2}^{2N-2} + \beta_{N-3} a_{2N-3}^{2N-2} \qquad (3.31)$$

$$a_{2N}^{2N} = \beta_{N-1} a_{2N-2}^{2N-2} \qquad (3.32)$$

The coefficients can be used for the implementation of the SPZC-based MNF using ARMA modeling. The performance of this filter appears to be significantly better then the error compensation MNF, as will be shown in the simulation section. Because the compensation section of the SPZC-based MNF has been obtained in a closed form by simply knowing the

notch frequency and the pole-contraction factor, this method becomes the more attractive one.

3.3.3 LMS-Optimal Roots of the MNF

The compensation poles have been constrained to be the same as the original CNF poles, the task at hand is essentially that of an FIR type filter design. For solving this problem, we modify an existing optimization technique [55] that can be used to directly yield the coefficient of the MA part of the compensation section.

Optimization methods for evaluating the filter coefficients involve three steps [56]. In the first step, an error function that depends on the difference between the desired and actual is formulated. In the second step this error function is minimized to yield the desired filter coefficients. In the third step an allpass phase equalizer is designed to achieve the desired phase. It is possible to incorporate the desired phase and amplitude response in the above step two, thus eliminating step three as in [55]. We modify the technique of [55] to suit our problem. Let us rewrite the MNF of (3.22) as

$$H'(z) = \frac{P(z)R(z)}{Q(z)} \tag{3.33}$$

where

$$\begin{aligned} P(z) &= 1 + p_1 z^{-1} + p_2 z^{-2} \\ R(z) &= 1 + r_1 z^{-1} + \cdots + r_{2N-2} z^{-2N+2} \\ Q(z) &= 1 + q_1 z^{-1} + \cdots + q_{2N} z^{-2N} \end{aligned}$$

The numerator of the MNF has been split into two parts: the known MA part $P(z)$ of the original notch filter ($\bar{B}(z)$ in (3.22)) and the unknown MA part $R(z)$ of the compensation section of the filter ($\tilde{B}_c(z)$ in (3.22)). $Q(z)$ is of course known by the earlier constraint on the pole locations and is equivalent to $\bar{A}(z)\tilde{A}_c(z)$. It is only necessary to determine $R(z)$. It is important to realize that we need not design the complete MA part, but only the compensation section. Importance of this procedure is realized in evaluating closed form expressions for the compensation filter coefficients.

The frequency response of (3.33) can be written as

$$H'(\omega) = \frac{(1 + p^T c_1(\omega) + jp^T s_1(\omega))(1 + r^T c_2(\omega) + jr^T s_2(\omega))}{1 + q^T c_3(\omega) + jq^T s_3(\omega)} \tag{3.34}$$

where

$$\begin{aligned} p &= [p_1 \quad p_2]^T \\ r &= [r_1 \quad r_2 \quad \cdots \quad r_{2N-2}]^T \\ q &= [q_1 \quad q_2 \quad \cdots \quad q_{2N}]^T \end{aligned}$$

and

$$\begin{aligned} c_1(\omega) &= [\cos\omega \quad \cos 2\omega]^T \\ s_1(\omega) &= [-\sin\omega \quad -\sin 2\omega]^T \\ c_2(\omega) &= [\cos\omega \quad \cos 2\omega \quad \cdots \quad \cos(2N-2)\omega]^T \\ s_2(\omega) &= [-\sin\omega \quad -\sin 2\omega \quad \cdots \quad -\sin(2N-2)\omega]^T \\ c_3(\omega) &= [\cos\omega \quad \cos 2\omega \quad \cdots \quad \cos 2N\omega]^T \\ s_3(\omega) &= [-\sin\omega \quad -\sin 2\omega \quad \cdots \quad -\sin 2N\omega]^T. \end{aligned}$$

By using the ideal frequency response of the notch filter, i.e.,

$$H_d(\omega) = \begin{cases} 1 & \omega \neq \omega_o \\ 0 & \omega = \omega_o \end{cases} . \tag{3.35}$$

as the desired response (where subscript 'd' denotes desired), the error function can now be formulated as

$$e(\omega) = H_d(\omega) - \frac{P(\omega)R(\omega)}{Q(\omega)} \tag{3.36}$$

Since the desired frequency response $H_d(\omega)$ is a real function for all ω, the weighted error function can be written as

$$\begin{aligned} e(\omega)Q(\omega) &= Q(\omega)H_d(\omega) - P(\omega)R(\omega) \\ &= e_r(\omega) + je_i(\omega) \end{aligned} \tag{3.37}$$

where

$$
\begin{aligned}
e_r(\omega) &= H_d(\omega) - 1 + H_d(\omega)q^T c_3(\omega) - r^T c_2(\omega) - p^T c_1(\omega) r^T c_2(\omega) + p^T s_1(\omega) r^T s_2(\omega) \\
e_i(\omega) &= H_d(\omega)q^T c_3(\omega) - r^T s_2(\omega) - p^T c_1(\omega) r^T s_2(\omega) - p^T s_1(\omega) - p^T s_1(\omega) r^T c_2(\omega)
\end{aligned}
$$

the subscripts 'r' and 'i' denote the real and imaginary parts, respectively. By using $e_r(\omega)$ and $e_i(\omega)$, the weighted mean square error is defined as

$$E_{mse} = \int_0^\pi \left[e_r^2(\omega) + e_i^2(\omega) \right] d\omega \tag{3.38}$$

and the FIR filter coefficients r_i's, $i = 1, 2, \ldots 2N-2$ can be determined by minimizing E_{mse} w..r.t. r, i.e., $dE_{mse}/dr = 0$. On performing this minimization, we obtain a system of linear equations

$$\frac{\partial E_{mse}}{\partial r} = \int_0^\pi Q'(\omega) r d\omega - \int_0^\pi d'(\omega) d\omega = 0 \tag{3.39}$$

where

$$
\begin{aligned}
Q'(\omega) &= \left[1 + 2p^T c_1(\omega) + p^T c_1(\omega) c_1^T(\omega) p + p^T s_1(\omega) s_1^T(\omega) p \right] \cdot \\
&\quad \left[c_2(\omega) c_2^T(\omega) + s_2(\omega) s_2^T(\omega) \right]
\end{aligned} \tag{3.40}
$$

and

$$
\begin{aligned}
d'(\omega) &= \left[1 + 2p^T c_1(\omega) + p^T c_1(\omega) c_1^T(\omega) p + p^T s_1(\omega) s_1^T(\omega) p \right] c_2(\omega) \\
&\quad - H_d(\omega) \left\{ c_2(\omega) + p^T c_1(\omega) c_2(\omega) - p^T s_1(\omega) s_2(\omega) \right. \\
&\quad + c_3^T(\omega) q \left[c_2(\omega) + P^T c_1(\omega) c_2(\omega) - P^T s_1(\omega) s_2(\omega) \right] \\
&\quad \left. + s_3^T(\omega) q \left[s_2(\omega) + p^T c_1(\omega) s_2(\omega) + p^T s_1(\omega) c_2(\omega) \right] \right\}.
\end{aligned} \tag{3.41}
$$

Evaluating the integral leads to the following system of linear equation that can easily be solved to yield the desired filter coefficients, **r**

$$Q\mathbf{r} = d \tag{3.42}$$

where

$$Q = \begin{bmatrix} (\theta^2+2) & 2\theta & 1 & & & & & & \\ 2\theta & (\theta^2+2) & 2\theta & 1 & & & & & \\ 1 & 2\theta & (\theta^2+2) & 2\theta & 1 & & & \mathbf{0} & \\ & & & \ddots & & & & & \\ & & & & 1 & 2\theta & (\theta^2+2) & 2\theta & 1 \\ & & \mathbf{0} & & & 1 & 2\theta & (\theta^2+2) & 2\theta \\ & & & & & & 1 & 2\theta & (\theta^2+2) \end{bmatrix}$$

and

$$\begin{aligned} d(1) &= -\theta + q_1 + \theta q_2 + q_3 \\ d(2) &= -1 + q_2 + \theta q_3 + q_4 \\ d(i) &= q_i + q_{i+1} + q_{i+2} \qquad for \;\; i = 3, 4, \ldots 2N-2. \end{aligned}$$

Since Q is a positive-definite matrix, a unique solution for the coefficient of MA part of the compensation section of MNF can always be found. The fact that Q is a band matrix can be exploited to efficiently evaluate $\mathbf{r}$.

In order to obtain closed-form expression for Q and d above, it is necessary to integrate over the entire frequency range $[0, \pi]$. Since $H_d(\omega)$ is unity everywhere except at $\omega = \omega_o$, we can set the desired filter response $H_d(\omega)$ to unity over the entire spectrum. For the purpose of integration this has no effect, since an integral over a single point has no contribution to the overall integration on the interval $[0, \pi]$. The fixed MA part (known part) of the MNF, $P(z)$, forces a null at $\omega = \omega_o$ and since the MNF poles are fixed, this procedure has no effect in the overall filter response.

To illustrate the design procedure we present the following example.

Example 3.2 A sixth order MNF is composed of a second order CNF and a fourth order compensation section,

$$H'(z) = \left(\frac{1 + \theta z^{-1} + z^{-2}}{1 + \alpha\theta z^{-1} + \alpha^2 z^{-2}} \right) \left(\frac{1 + b_1^c z^{-1} + \cdots + b_4^c z^{-4}}{1 + a_1^c z^{-1} + \cdots + a_4^c z^{-4}} \right) \tag{3.43}$$

where a_i^c, $i = 1,2,3,4$ are known since the poles are constrained as described earlier and b_i^c, $i = 1,2,3,4$ are obtained from the following system of linear equations

$$\begin{bmatrix} \theta^2+2 & 2\theta & 1 & 0 \\ 2\theta & \theta^2+2 & 2\theta & 1 \\ 1 & 2\theta & \theta^2+2 & 2\theta \\ 0 & 1 & 2\theta & \theta^2+2 \end{bmatrix} b^c = \begin{bmatrix} -2\theta - 1 + q_1 + \theta q_2 + q_3 \\ q_2 + q_3 + q_4 \\ q_3 + q_4 + q_5 \\ q_4 + q_5 + q_6 \end{bmatrix} \tag{3.44}$$

and the q_i's are the coefficients of the AR part of the total MNF that can be obtained by multiplying out the denominator of (3.13). Figure 3.7 shows the distribution of the poles and zeros for an MNF with notch frequency $\omega_o = 1$ (i.e. $\theta = -2\cos\omega_o = -1.0806$) and Table 3.1 shows the resulting coefficients.

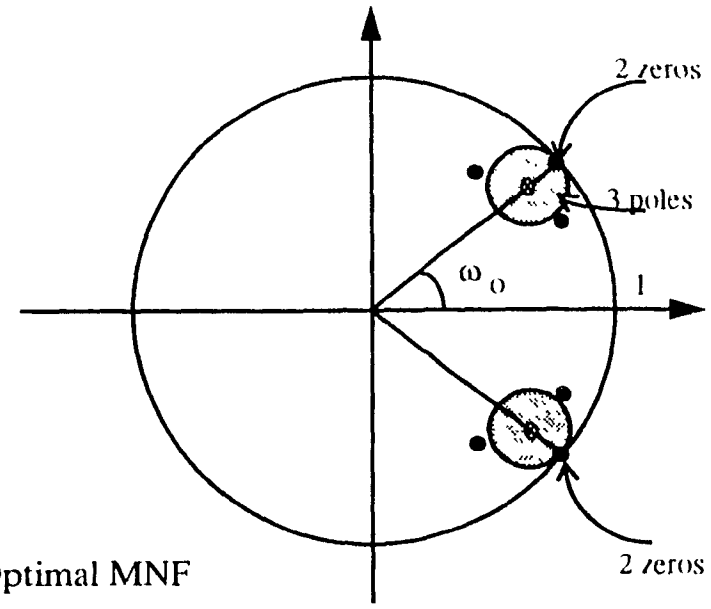


Figure 3.7: Pole/zero constellation of a sixth order Optimal-MNF

	i=0	i=1	i=2	i=3	i=4	i=5	i=6
p_i	1	-1.0806	1				
r_i	1	-1.5055	1.5434	-.7907	0.2593		
q_i	1	-2.5935	4.1620	-3.9657	2.6637	-1.0623	0.2621

Table 3.1: Coefficients of the two MA parts and the AR part of the sixth order MNF

3.4 Simulations

The proposed methods are more complex in design than the standard second order CNFs. The increase in complexity, however, is minimal, since simple closed-form solutions are presented. The issue of complexity need not be of any great concern since the design of MNFs is done off-line as it would be done for the standard CNF.

In this section, we evaluate the performance of the MNFs designed via the three proposed techniques: the Error-Correction Approach (ECA), the Symmetric Pole/Zero Constellation (SPZC) and the LMS-Optimal (OPT) approach. We compare the MNFs, so designed, against standard second order notch filter of (3.17). In comparing the performance of these filters we will use the notion of transparency and transient durations as described earlier. In addition to these measures we will also use the ideas of notch bandwidth (NBW), stopband and passband errors. In the following, the LMS-Optimal, SPZC and the ECA MNFs will be denoted as $H'_{opt}(z)$, $H'_{spzc}(z)$ and $H'_{eca}(z)$, respectively, while the standard CNF is denoted as $H(z)$.

It is clear that as the pole-contraction factor α increases, the transient duration also increases [53]. Similarly, for a fixed α as the order of the filter increases, the transient duration also increases. Thus, in comparing these filters, the value of α is adjusted such that all filters have the same transient duration for the above performance measures to be compared on an equal basis.

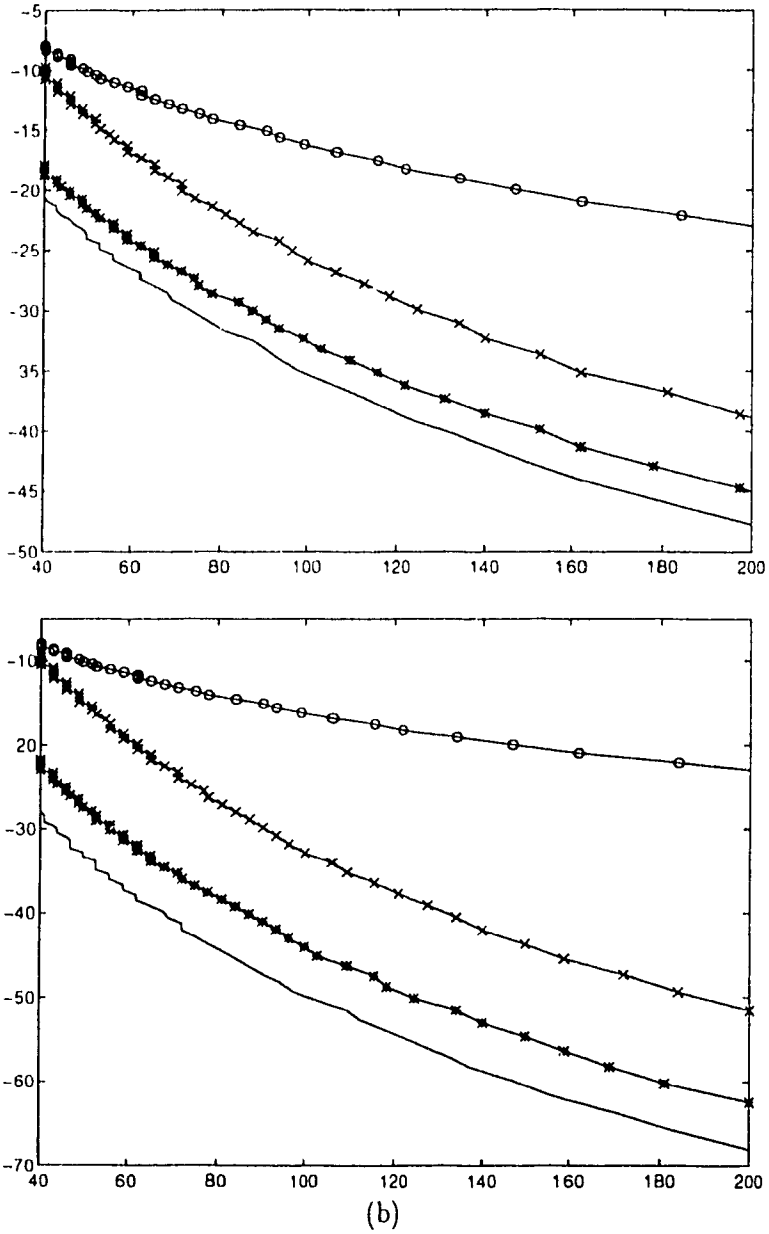
Figure 3.8a show the transparency misadjustment coefficient Γ as a function of the transient duration τ for a conventional second order notch filter, $H(z)$ and fourth order MNFs designed via the three proposed approaches, i.e. $H'_{eca}(z)$, $H'_{spzc}(z)$ and $H'_{opt}(z)$. Clearly, $H(z)$ exhibits the worst transparency and the $H'_{opt}(z)$ the best transparency; while the $H'_{spzc}(z)$ and $H'_{eca}(z)$ are somewhere in between the two. The transparency coefficient for $H'_{eca}(z)$ approaches that of $H(z)$ for lower quality filters (lower values of α or equivalently those with shorter transient durations), thus, showing no improvements over $H(z)$. Even as we increase the order of $H'_{eca}(z)$ to six in Figure 3.8b, there is no improvement for lower quality filters. The transparency, however, for the $H'_{spzc}(z)$ shows a significant improvement over $H(z)$. From the given curves in Figures 3.8a and b, it is apparent that the transparency curves of the $H'_{opt}(z)$ are approximately 3 and 6 dB

better than the SPZC-MNFs for the fourth and sixth order filters, respectively. The transparency performance of the SPZC-MNFs is very close to the optimal MNFs. As the quality of the filter is increased, the improvements for all MNFs over the conventional second order filter are seen to also increase. Figure 3.9a compares the stopband error curves for fourth order MNFs against the conventional CNF. All three of the MNFs provide virtually identical performance which are approximately 2 dB better than that of the CNF. For lower quality filters the $H'_{eca}(z)$ is seen to approach the performance of CNF. In Figure 3.9b no significant improvement is observable by increasing the order of the MNFs to six. Figure 3.10a compares the stopband error curves for the fourth order MNFs against the conventional CNF. All three of the MNFs provide virtually identical performance which is approximately 2 dB better than that of the CNF. In Figure 3.10b by increasing the MNF order to six, an improvement of about 1 dB is observed. The improvement in the passband and the stopband appear to be insignificant at first glance. But, due to the averaging effect of the L_2-norm definition of these error, the very significant improvement away from the notch points are masked.

Figure 3.11a shows the normalized notch bandwidth for the CNF and the three proposed fourth order MNFs. All three of the MNFs have approximately the same notch bandwidth that is significantly lower than that of the CNF. As the quality of the filter is increased this improvement is reduced. This is expected as the pole contraction factor approaches unity, the errors in the CNF become negligible and the relative improvement becomes smaller. Figure 3.11b shows the results of increasing the order of the MNF to six. For low quality filter the notch bandwidth is seen to further decrease.

The following example illustrates the performance of the MNFs under the condition that the *transient duration is taken as a fixed parameter*, i.e., a vertical slice of Figures 3.8, 3.10, 3.9 and 3.11.

Example 3.3 In this example, we design sixth order MNFs with a transient duration of approximately 56 samples. We adjust α for $H(z)$ such that it has the same quality as the MNFs (quality referring to the transient durations). Table 3.2 shows the zero locations of the four filters (where ω_o is the normalized notch frequency in the range $[0, \pi]$), which are also shown as a pole/zero plot in Figure 3.12. In Figure 3.12, the zeros are indeed



(b)

Figure 3.8: Transparency misadjustment error vs. transient duration. o = CNF; x = ECA-MNF; * = SPZC-MNF; - = OPT-MNF.

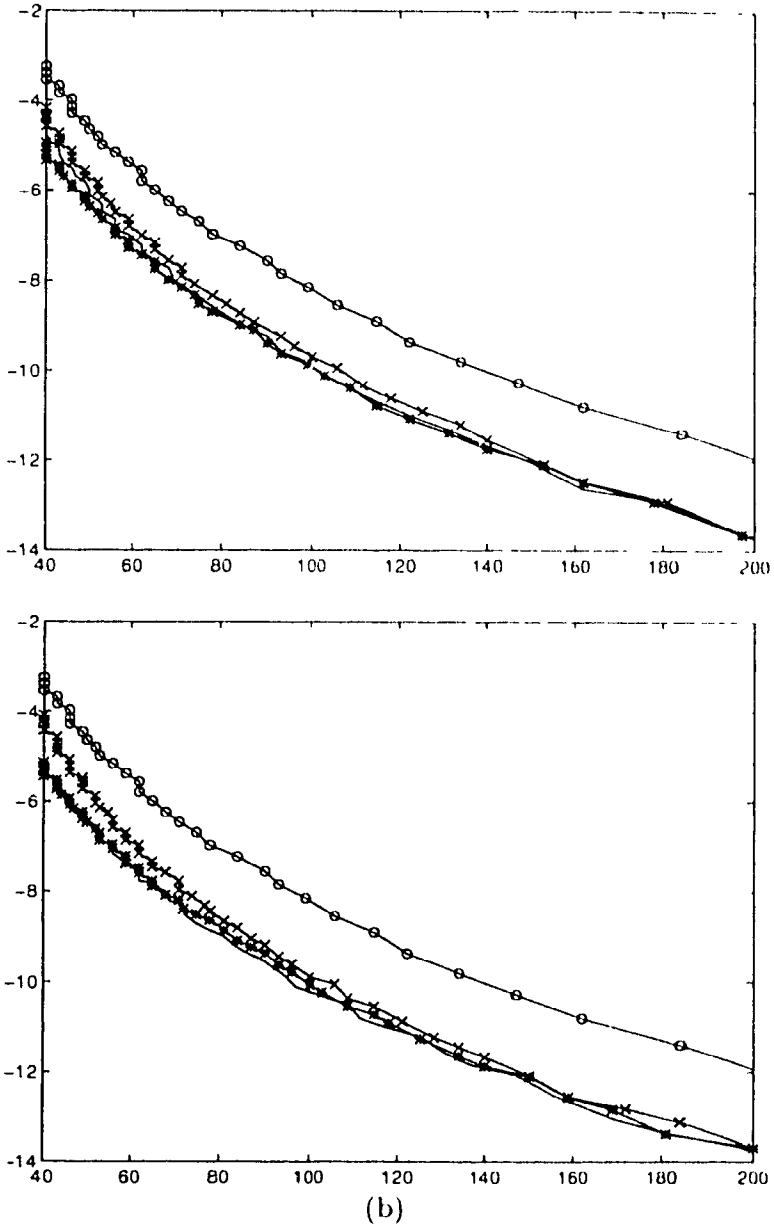


Figure 3.9: Stopband *mse* vs. transient duration o = CNF; x = ECA-MNF; * = SPZC-MNF; – = OPT-MNF.

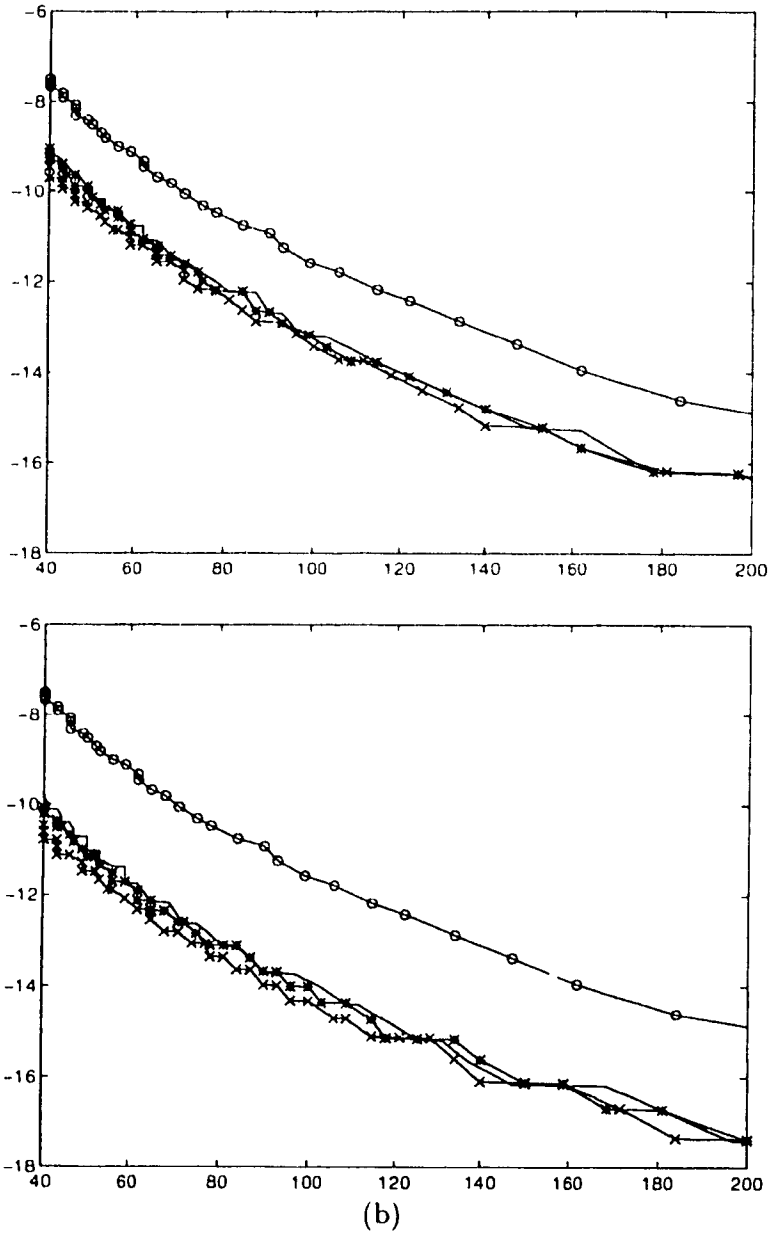


Figure 3.10: Passband *mse* vs. transient duration. o = CNF; x = ECA-MNF; * = SPZC-MNF; - = OPT-MNF.

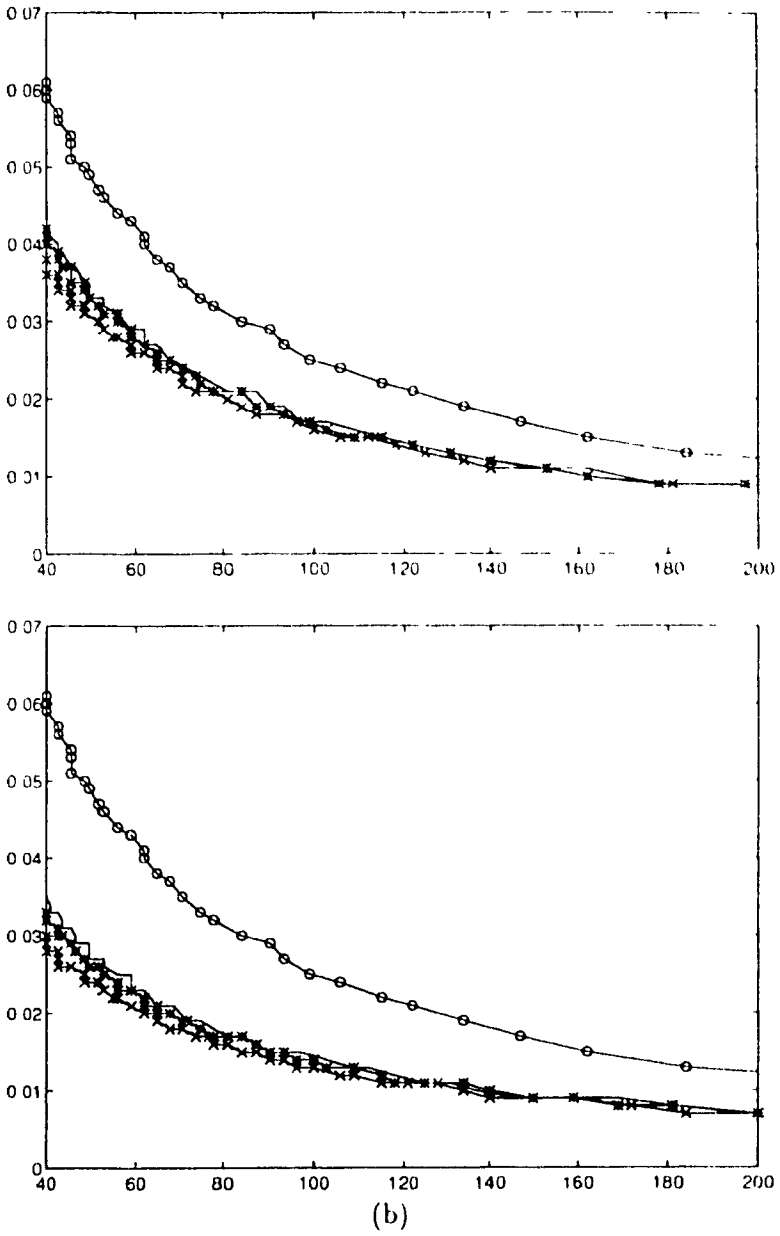


Figure 3.11: Notch bandwidth vs. transient duration. o = CNF; x = ECA-MNF; * = SPZC-MNF, - = OPT-MNF.

clustered around a circle of radius $(1-\alpha)$ centered at the poles as discussed previously. The various measures of performance are listed in Table 3.3 where the transient τ is measured in number of samples and the notch bandwidth (NBW) shows the % reduction relative to that of $H(z)$.

$H(z)$	$H'_{opt}(z)$	$H'_{spz}(z)$	H'_{eca}
$e^{\pm j\omega_o}$	$e^{\pm j\omega_o}$	$e^{\pm j\omega_o}$	$e^{\pm j\omega_o}$
-	$0.7490e^{\pm j(\omega_o+0.2019)}$	$0.7186e^{\pm j(\omega_o+0.2459)}$	$0.7128e^{\pm j(\omega_o+0.3416)}$
-	$0.6825e^{\pm j(\omega_o+0.2183)}$	$0.7186e^{\pm j(\omega_o+0.2459)}$	$0.7666e^{\pm j(\omega_o+0.2903)}$

Table 3.2: Zero locations of the three filters considered

	$H(z)$	$H'_{opt}(z)$	$H'_{spzc}(z)$	$H'_{eca}(z)$
α	0.8730	0.8010	0.7980	0.7920
τ (samples)	56	56	56	56
Stopband error (dB)	-5.2721	-7.19073	-7.1442	-6.6107
Passband error (dB)	-9.1009	-11.3713	-11.7053	-11.8827
Γ dB	-11.2253	-36.1316	-30.4725	-18.3038
NBW		11.86%	46.51%	48.84%

Table 3.3. Performance of sixth order MNF vs. NF for fixed transient duration

From Table 3.3 it is seen that the performance of the three MNFs is virtually identical. Clearly, all three MNFs are superior to the CNF in all respects. The $H'_{opt}(z)$ has the best transparency misadjustment coefficient and the notch bandwidth of the $H'_{eca}(z)$ is only marginally better that the other two MNFs.

Although the MSE is a good indicator of the passband and stopband errors, it fails to adequately describe the frequency response of the notch filters away from the notch point. The dynamic range of the magnitude response $1 - H(\omega)$ varies from around unity near the notch point to very small values increasingly away from this point. Figure 3.13 shows this error ($\mid 1 - H(\omega) \mid$) as a function of frequency. It is evident that increasingly away from the notch point, the errors in the magnitude response are greatly minimized, leading to am improved allpass performance.

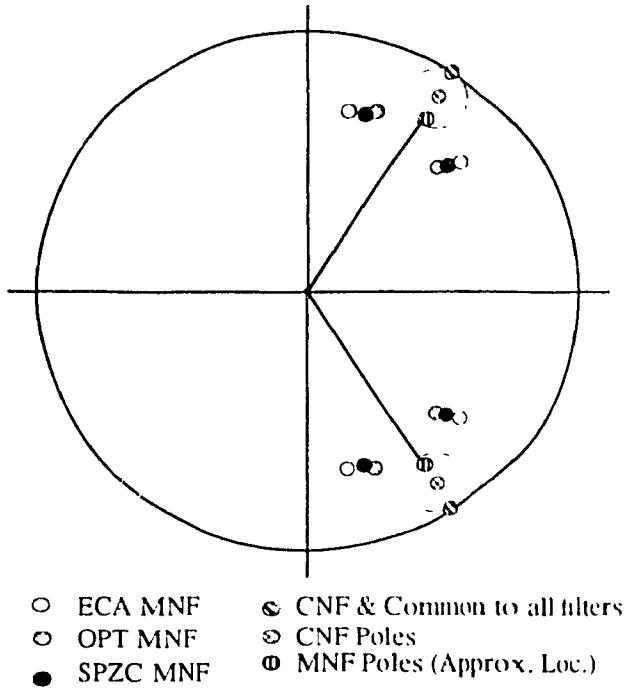


Figure 3.12: Pole/zero locations of sixth order MNFs and a CNF

Example 3.4 In this example, we show that as a consequence of the improvements in the MNFs, it is possible to achieve reduced transient durations. We reduce the value of α for each of the MNFs such that the error in the frequency response is equivalent to that of the conventional second order CNF. Here, for the purpose of illustration, we use the notch bandwidth (NBW) as the error criterion. The notch bandwidth is often an important criterion in the design of NFs - for example, in the separation of closely spaced sinusoids. In adaptive NFs (CANFs) [26], the accuracy of the estimated coefficients is directly dependent on the NBW, i.e., the narrower the the NBW, the more accurate are the estimates of filter coefficients. Of course, this choice of error criterion should be application dependent – it may be desirable to consider the passband errors or the transparency misadjustment in a line-enhancement application.

The results are listed in Table 3.4, where the normalized NBWs are equal for all cases. The $H'_{opt}(z)$ and $H'_{spzc}(z)$ show approximately 45% reduction in transient duration and $H'_{eca}(z)$ shows a 56% improvement. The greater improvement in the ECA-MNF is at the cost of degradation in the performance of the stopband and the all important

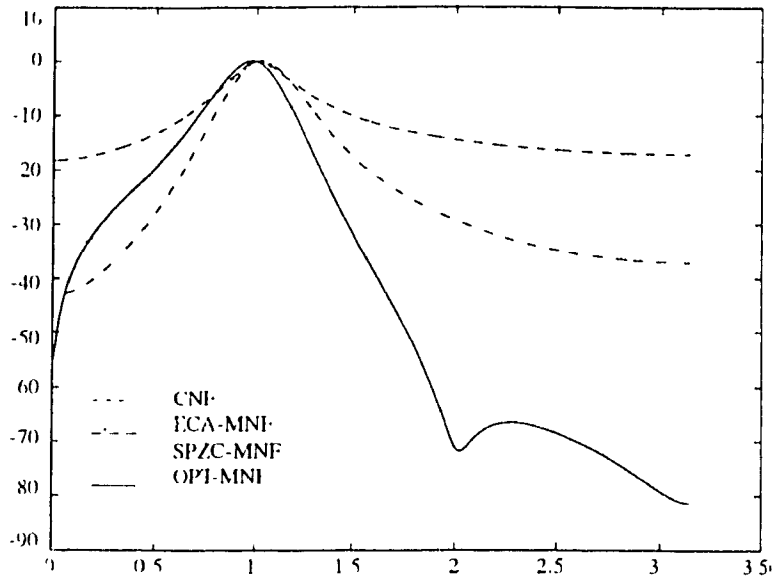


Figure 3.13: Error in the magnitude response as a function of frequency.

transparency misadjustment coefficient, Γ. For the other two MNFs the reduction in the transient duration does not force the other measures to be significantly reduced. In fact, the transparency misadjustment coefficient still shows large improvements.

	$H(z)$	$H'_{opt}(z)$	$H'_{spzc}(z)$	$H'_{eca}(z)$
α	0.9000	0.7170	0.6965	0.6225
τ (samples)	71	40	37	31
Stopband error (dB)	-6.4534	-5.1528	-5.0284	-2.2272
Passband error (dB)	-10.0449	-9.8408	-9.8063	-9.7673
Γ dB	-13.2078	-28.1473	-21.1084	-4.4168

Table 3.4: Performance of sixth order MNFs vs. NF for fixed notch bandwidths.

Example 3.5 This example is similar to the last example. Here, instead of using the NBW as an error criterion we use the passband errors. The results are listed in Table 3.5. The results show a similar trend as that of the last example.

	$H(z)$	$H'_{opt}(z)$	$H'_{spzc}(z)$	$H'_{eca}(z)$
α	0.9000	0.7270	0.7040	0.6490
τ (samples)	71	41	40	31
Stopband error (dB)	-6.4534	-5.3685	-5.1562	-2.9022
Passband error (dB)	-10.0119	-9.9935	-9.9367	-10.0358
Γ dB	-13.2078	-28.9426	-21.6726	-6.0504
NBW	-	2.86%	2.86%	7.71%

Table 3.5: Performance of sixth order MNFs vs. NF for fixed passband errors

3.5 Conclusion

In practical situations the values of the pole-contraction factor α, in the range of .8 to .99, cause phase and magnitude response distortions in the NF. It has been shown, that these deviations in the frequency response can be combated or reduced if the order of the CNF is increased by strategic placement of additional pole/zero pairs. For a given transient duration, this significantly improves the transparency property of the new filter (MNF). Due to the transparency/transient duration tradeoff, this improvement in transparency can be traded for a reduced transient duration. Simulation results have been used to show that significant improvement in the transient duration that can be achieved with this tradeoff.

Three methods of evaluating the compensation filter roots have been presented, namely the Error-Correction Approach (ECA), Symmetric Pole/Zero Compensation (SPZC) and the LMS-Optimal. The first two of these present closed-form suboptimal methods of evaluating the compensation roots. The third method is optimal in the LMS sense, for which simple analytic solution has been presented. Due to the optimality, this method also serves as a lower bound for the errors corresponding to the suboptimal methods. Simulation results indicate that the LMS-Optimal MNF is negligibly better than the SPZC

methods, while both are significantly better than the ECA. The LMS-Optimal method allows for direct evaluation of the filter coefficients that simplifies and formalizes the design of MNFs used for the analysis of short record signals. For this reason it is the preferable one

Chapter 4

Statistically Optimal Null Filters (SONF)

In this chapter, we propose an alternative nonparametric statistically optimal method of null filtering. One of the important aspects of this new method lies in its ability to process signals of short record lengths.

Synthesis of optimum linear filters involves choosing the system in such a way as to satisfy certain rules that make it optimal. In this design process three aspects must be addressed: the *input specification*, *system constraints* and the *criterion of optimality*. Input specification implies that some *a priori* knowledge of the input signal is available. System constraints define the type of resulting system. For example, if the input consists of time-varying signal in noise, then we may want to design a linear time-varying (LTV) system. Criterion of optimality reflects a meaningful measure of the goodness as related to the problem at hand. For example, we may choose Minimum Mean-Square Error (MMSE) or the maximum output Signal to Noise Ratio (SNR) [57]. In this chapter, we exploit both of these measures.

There exist two major problems in signal processing: signal separation (enhancement/suppression) and signal detection. For signal separation, it is customary to use the MMSE criterion to describe the performance of processing. In stationary environments, it is possible to design a linear time invariant (LTI) system that is optimal in the MMSE sense. Such a filter is known as the Weiner filter [58] and its design is based on the knowledge of the signal and the noise power spectral densities. As for the detection problem, filtering is based on the knowledge of the signal shape (this knowledge may or may not be complete) and the optimality criterion is the maximum output SNR (SNR_o). Such a filter is known as the Matched Filter (MF) [58, 59, 60].

We propose to combine the MF (designed for detection) and the least-squares (LS) optimization approach to solve the problem of enhancement/suppression of short-duration signal embedded in noise. It is interesting to note that MFs are used exclusively for detection purposes, most notably in communications applications, see for example [61, 62]. In spite of this fact, in our approach, we propose to use them as a fundamental building block for the *filtering* of short-duration signals.

Note that designing a filter to estimate a signal, $s(t)$, can be considered to be equivalent to designing a null filter. That is, if an optimum linear filter $H(\omega)$ can be designed to estimate $s(t)$ in the presence of interference, then the null filter is simply defined as

$1 - H(\omega)$ and vice versa. In this sense, designing a null filter or a line enhancer is equivalent.

Section 4.1 presents the statement of problem. Some definitions and ideas that will be used in the subsequent sections are also described. In Section 4.2, we exploit the idea of MFs and introduce the Instantaneous Matched Filters (IMFs). In Section 4.3, we use the IMF as a basic building block and present the new approach of SONFs. Several kinds of the SONF are presented. Globally optimal SONF is investigated in Section 4.4. The globally optimal SONF cannot be easily implemented, as such, implementation issues are considered in Section 4.5 and a discrete-time SONF is developed. A recursive SONF is also introduced, wherein reduced *a priori* information is required. Section 4.6 contains some analysis verifying the fact that under some constraint the SONFs may be considered as a new implementation of the Kalman Filter. In Section 4.7 we develop an adaptive estimation method of the damping coefficients of damped sinusoids using the SONF.

4.1 Preliminaries

In this section we present the statement of the problem in the context of processing transient signals. Definition of transients, as considered in this thesis, is given. We summarize the results of Optimum Linear Filter as presented in [11]. The idea of representation of signal with an orthogonal basis set is reviewed. In so doing, in the framework of on line estimation, we consider the idea of the classical Gram-Schmidt orthogonalization procedure for arbitrary or dynamically changing time intervals.

4.1.1 Problem Statement

A continuous-time signal $x'(t)$ is measured on an observation interval of T_{ob} seconds. Most of the time, the measured signal $x'(t)$ consists of only noise. Occasionally, it contains an additive mixture of signal $s(t)$ and noise $n(t)$. The signal $s(t)$, when present, has a short duration as compared to the observation interval. It is often of the decaying type, that is, its amplitude decays as time increases. Moreover, $s(t)$ may consist of an additive mixture of several such signals. Signals of this form are often called transients. Very often the terms "short or finite duration" and "transients" are used synonymously.

In general, the problem of estimating transient signal waveforms can be formulated in two parts:

- Detection of the presence of the transient and the estimation of the *time-of-arrival*, TOA, (see Chapter 1 for the definition of TOA).
- Estimation of the waveform given the TOA.

In our work we are concerned with the second of the two parts, that is, given the presence of the transient and its TOA, to estimate the signal waveform. Of course, some *a priori* knowledge of the transient waveform is also available. This does not reduce the generality of the problem as techniques exist for transient detection [8, 63, 64, 65] and TOA estimation [6].

To formulate the problem at hand, consider a received or measured signal

$$x'(t) = s(t) + n(t) \qquad t \in [0, T_{ob}] \tag{4.1}$$

where $s(t)$ represents the transient (message or the desired signal) that may be a sample of a random process, $n(t)$ is additive white gaussian noise of zero-mean that is statistically independent of $s(t)$, and $[0, T_{ob}]$ is the observation interval. It is in knowing that the signal $s(t)$ is nonvanishing in the observation interval - vanishing only for a subset of the interval $[0, T_{ob}]$, i.e., not existing from 0 to TOA - and given the TOA, that we may truncate that portion of the received waveform prior to the existence of $s(t)$. Using a new observation interval T corresponding to the time from TOA of $s(t)$ to T_{ob}, we can rewrite (4.1) as

$$x(t) = s(t) + n(t) \qquad t \in [0, T] \tag{4.2}$$

where $t = 0$ now corresponds to the TOA of $s(t)$. The signal $s(t)$ may vanish before T_{ob} is reached. For this time duration, we consider $s(t)$ to be of zero value.

We treat the general case where $s(t)$ is a random signal that can be represented as

$$s(t) = \sum_{i=1}^{N} v_i \phi_i(t) \qquad t \in [0, T] \tag{4.3}$$

where v_i's are the unknown random variables and $\phi_i(t)$'s are the known set of basis functions.

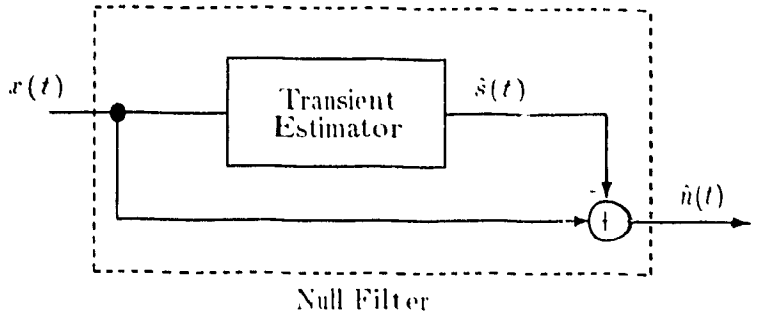


Figure 4.1: Suppression of Transient Signal

The problem at hand can be defined as the estimation of the signal $s(t)$ under the MSE optimality criterion and thus to implement null filtering by subtracting $\hat{s}(t)$ (the estimate of $s(t)$) from the input mixture as in Figure 4.1.

4.1.2 Optimum Linear Filter

The problem of optimally estimating a message signal in the presence of interfering noise was considered by Van Trees in [44]. Here, we briefly summarize these results for a later comparison.

System of interest is shown in Figure 4.2. The received or observed signal consists of an additive mixture of the desired signal and noise as in (4.2). The signal $s(t)$ is, in general, a sample of a random process with covariance function $K_s(t,u)$ and $n(t)$ is the zero-mean white gaussian noise with covariance $K_n(t,u) = N_o\delta(t-u)$. The noise is statistically independent of $s(t)$. Since $s(t)$ is not necessarily stationary, it is anticipated

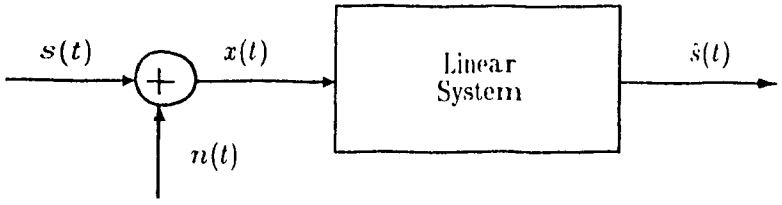


Figure 4.2: Linear estimator of signal $s(t)$.

that the optimal solution may require a time-varying filter, $h(t,u)$ output at time t given input at time u. For a realizable $h(t,u)$, the output due to the input $x(u)$ is written as a

convolution integral,

$$\hat{s}(t) = \int_0^T h(t,u)x(u)du. \tag{4.4}$$

The objective is to find $h(t,u)$ such that $\hat{s}(t)$ represents $s(t)$ with a minimum mean-square error over the observation interval $[0,T]$.

A solution to $h(t,u)$ is possible using standard variational techniques; however, Van Trees [41] uses a less formal but lengthy approach to yield the following optimal solution

$$h_o(t,u) = \sum_{i=1}^{\infty} \frac{\mu_i}{\mu_i + N_o} \phi_i(t)\phi_i(u) \tag{4.5}$$

in terms of μ_i's (eigenvalues) and $\phi_i(t)$'s (eigenfunctions) of $K_s(t,u)$. The subscript 'o' in $h(t,u)$ of (4.5) denotes optimal. In general, the number of eigenvalues is large and (4.5) represents a practical solution for cases where the number of significant eigenvalues is small. Figure 4.3 shows a k-term implementation. The mean-square error in the estimation of $s(t)$ is expressed as

$$\zeta(t) = \frac{N_o}{2} \sum_{i=1}^{k} \frac{\mu_i}{\mu_i + N_o/2} \phi^2(t), \qquad t \in [0,T]. \tag{4.6}$$

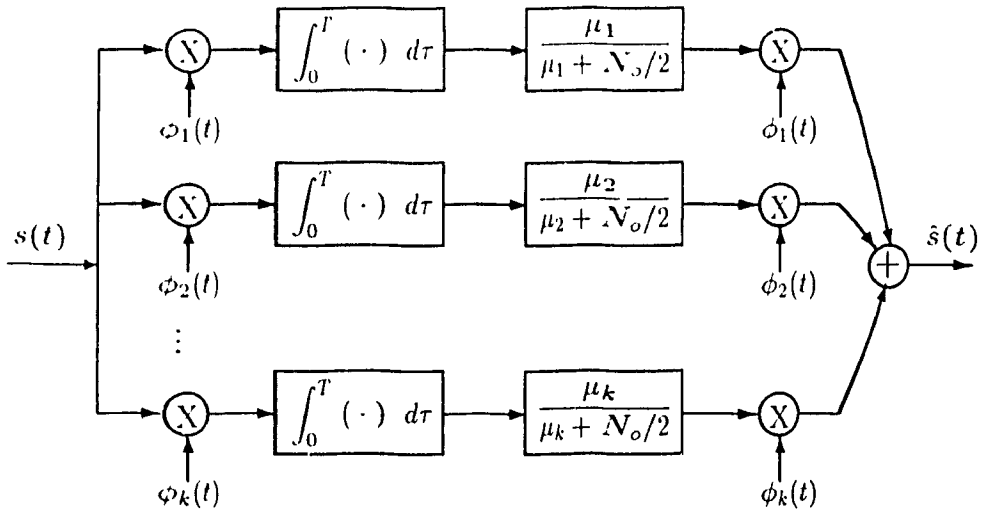


Figure 4.3: k-term Linear Optimum Filter

4.1.3 Interval of Orthogonality

In many applications such as speech and image coding, orthogonal basis functions, like those in the Fourier Transform and the Discrete Fourier Transform (DFT), are used to reduce the correlation in the data. Recently, the use of wavelets has been suggested for this purpose [66].

Often Fourier Series is used as a close approximation for nonperiodic processes. For such processes the Fourier coefficients are generally correlated [67]. Correlation, however, can be reduced by increasing the observation interval T. If the covariance function is not very smooth, the decay in correlation can be slow. For example, the correlation in the Fourier coefficients of a signal with an exponential covariance function $K_s(t,u) = e^{-|t-u|}$ decays at a rate of $1/T^2$ as T increases, which is slow compared to the exponential decay of $K_s(t,u)$ [68]. Furthermore, increasing T indefinitely may render the calculations unstable creating overflow problems. Of course, for short data or transient problems this approach is clearly not feasible.

None of the above techniques reduce the correlation in the coefficient of the expansion as do the eigenfunctions. A signal $s(t)$ with a well defined covariance function $K_s(t,u)$ over an interval $[0,T]$ can be represented as a series expansion

$$s(t) = \sum_{i=1}^{N} v_i \phi_i(t) \tag{4.7}$$

where $\{\phi_i(t), i = 1, 2, \ldots, N\}$ is a set of orthonormal functions. This set is the solution of integral eigen equation

$$\int_0^T K_s(t,u)\phi_i(t)du = \mu_i \phi_i(t). \tag{4.8}$$

The significance of the above, known as the Karhunen-Loeve (KL) expansion, is that it provides an exact (in the mean-square error sense) whitened discretization of a continuous time process which yields a much simpler task of analyzing the process [14]. More often than not, such an expansion is only a theoretical concept since practical solutions to the integral eigen equation (4.8) are generally not possible.

Given an observation interval $[0,T]$, the KL expansion holds only over this interval. Over a new observation interval $[0,T']$, where $T' \neq T$, a new set of basis functions needs to be found. In another words, the set of basis functions over the observation interval $[0,T]$

is different from that over the observation interval $[0, T]$.

In an on-line estimation problem, as each new portion of the signal is received, the old set of orthogonal basis functions is not valid and a new one must be calculated. Alternatively, one may try to find a basis set that is orthogonal over an arbitrary interval $[0, t]$ where $t \leq T$, clearly an impossible task. One method of overcoming this difficulty is to orthogonalize the functions $\phi_i(t)$'s for each new time interval $[0, t]$ as t increases. We propose to use a sliding Gram-Schmidt (GS) orthogonalization procedure to achieve this. This procedure is different from the general GS orthogonalization procedure in that we do not orthogonalize over a fixed interval T, but rather at each instant of time over an interval that is dynamically increasing. Hence, the sliding aspect in this procedure.

The following illustrates the sliding GS procedure. Given,

$$s(t) = \sum_{i=1}^{N} c_i \phi_i(t) \qquad t \in [0, T], \tag{4.9}$$

where the set $\{\phi_i(t), i = 1, 2, \ldots, N\}$ may or may not be orthogonal over the interval $[0, T]$. We can formulate the following set of orthogonal functions over an arbitrary interval $t \in [0, T]$,

$$\begin{aligned}
\psi_1(t) &= \phi_1(t) \\
\psi_2(t) &= \phi_2(t) + \rho_{21}(t)\psi_1(t) \\
\psi_3(t) &= \phi_3(t) + \rho_{32}(t)\psi_2(t) + \rho_{31}(t)\psi_1(t) \\
&\vdots \\
\psi_N(t) &= \phi_N(t) + \sum_{i=1}^{N-1} \rho_{Ni}(t)\psi_i(t)
\end{aligned} \tag{4.10}$$

where

$$\rho_{ij}(t) = -\frac{\int_0^t \phi_i(\tau)\phi_j(\tau)d\tau}{\int_0^t \phi_j^2(\tau)d\tau} = -\frac{\langle \phi_i(t), \phi_j(t) \rangle_t}{\|\phi_j(t)\|_t^2} \quad for \quad \begin{cases} i &= 2, 3, \ldots, N \\ j &= 1, 2, \ldots, N-1 \end{cases} \tag{4.11}$$

For any given $t \in [0, T]$, the inner product $\langle \psi_i(t), \psi_j(t) \rangle_t = 0$ for $i \neq j$. The signal $s(t)$ can now be written in terms of set of functions, $\{\psi_i(t), i = 1, 2, \ldots N\}$, that are orthogonal

over any interval 0 to $t \in [0, T]$,

$$\begin{aligned} s(t) &= [v_1 - v_2\rho_{21}(t) - \cdots - v_N\rho_{N1}(t)]\, v_1(t) \\ &\quad + [v_2 - v_3\rho_{32}(t) - \cdots - v_N\rho_{N2}(t)]\, v_2(t) + \cdots + v_N v_N(t) \qquad (4.12) \\ &= A_1(t)v_1(t) + A_2(t)v_2(t) + \cdots + A_N(t)v_N(t) \qquad (4.13) \end{aligned}$$

By using the sliding GS orthogonalization procedure, it can be seen that we have transformed the constant coefficient signal expansion into one where the weighting coefficients are time-varying.

4.2 Instantaneous Matched Filter

In this section, we define the idea of Instantaneous Matched Filter. It is shown how the idea of MF can be extended to generate an output waveform that, in some ways, represents the message signal. In the subsequent sections, it will be shown how the resulting output of the IMF is further processed to extract the desired signal. The IMF forms a key building block in the proposed approach.

Consider the received signal $x(t)$ of (4.2) where the message signal $s(t)$ is completely represented by a single term ($N = 1$ in (4.3)) i.e. $s(t) = v_1\phi(t)$. For a specific time instant $t = t_I$ in the interval $[0, T]$, the observation interval becomes t_I. As time progresses, the interval of observation is continually increasing to the final value of T. If a matched filter (MF) is used to detect the signal at any given time, t_I, then at the output we obtain a signal that provides the maximum output signal to noise ratio, SNR_o, for the time interval $[0, t_I]$. Because the time interval or the frame of observation is continually increasing, at each considered time instant, the MF provides a new output signal and a new SNR_o. Hence, we have termed it the "Instantaneous Matched Filter (IMF)". Each new SNR_o reaches a maximum value for the considered time interval. Figure 4.4 depicts the IMF. Note that "Instantaneous" refers to the current time interval and not to the speed of processing via MFs. The output of an IMF at time t with $x(t)$ as the input is $\nu(t)$

$$\begin{aligned} \nu(t) = \int_0^t x(\tau)\phi(\tau)d\tau &= v_c(t) + n'_o(t) \\ &= s_o(t) + n'_o(t) \qquad (4.14) \end{aligned}$$

where $n'_o(t)$ represents the output noise, $s_o(t) = vc(t)$ is output signal that represents the message signal and $c(t)$ is defined as the L_2-norm of $\phi(t)$ over the observation interval t,

$$c(t) = \int_0^t \phi^2(\tau)d\tau = \| \phi(t) \|_t^2. \tag{4.15}$$

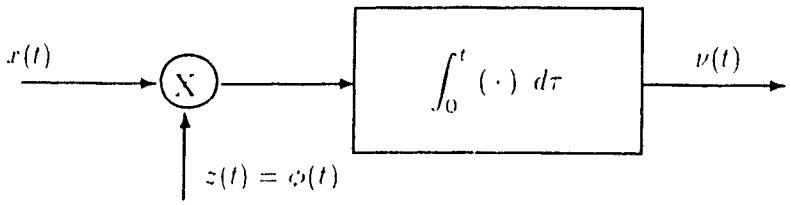


Figure 4.4: Instantaneous Matched Filter

Note that the output $\nu(t)$ of the IMF provides the most reliable (optimal) detection of $s(t)$ at each time t in the presence of zero-mean white gaussian noise [58, 59]. The only difference is that in the conventional MF, the upper limit of integration is a fixed moment T corresponding to the time (after observing the complete signal) at which the detection is made. In our case the same limit is an independent variable – the instantaneous time. *The IMF provides at each instant of time the maximum SNR_o at the output, independent of amplitude of the signal $s(t)$.*

$$\max \quad SNR_o(t) \quad \forall \quad v \in R \quad t \in [0, T]$$

4.3 Statistically-Optimal Null Filtering

In this section we describe our approach to optimal null filtering based on a combination of the maximum output SNR and the MSE criterion. An important consequence of such a formulation is that it enables us to process signals of short-duration. We first describe a simple coherent case and later generalize it to include the noncoherent case.

4.3.1 Coherent Null Filtering

Consider a signal where its waveform shape is *a priori* known. Let $s(t)$ in (4.2) be described as $s(t) = v\phi(t)$ where $\phi(t)$ is known and the amplitude v is an unknown random variable. Now, if we use the IMF of previous section then at time $t_1 \in [0, T]$, the output $\nu(t)$

(see (4.14)) provides the best measure for signal $s(t - t_I)$ to be detected in terms of SNR_o. The notation $s(t = t_I)$ denotes the signal $s(t)$ up to time $t_I \leq T$. Intuitively, one can understand that the maximum SNR_o at $t = t_I$ provides an increased measure (in terms of power) of the signal while reducing that of the noise. If this is done for every instant of time in $[0, T]$, then $\nu(t)$ provides a better measure or indication of the $s(t)$ than does $x(t)$. In light of the extensive literature on the detection of signal via MF this is quite obvious. Figure 4.5a shows the components of the output of the IMF (4.14), where the input desired signal was set to a sinusoidal waveform $s(t) = A\cos(\omega t)$ and the noise was unit-variance AWGN. Notice the envelope of $\nu(t)$ reflects the desired signal. Figure 4.5b shows the input and the IMF output SNRs. As time progresses, SNR_o provides a significantly better detection criterion for $s(t)$ than does SNR_i. We can exploit this enhanced criterion for the estimation of $s(t)$.

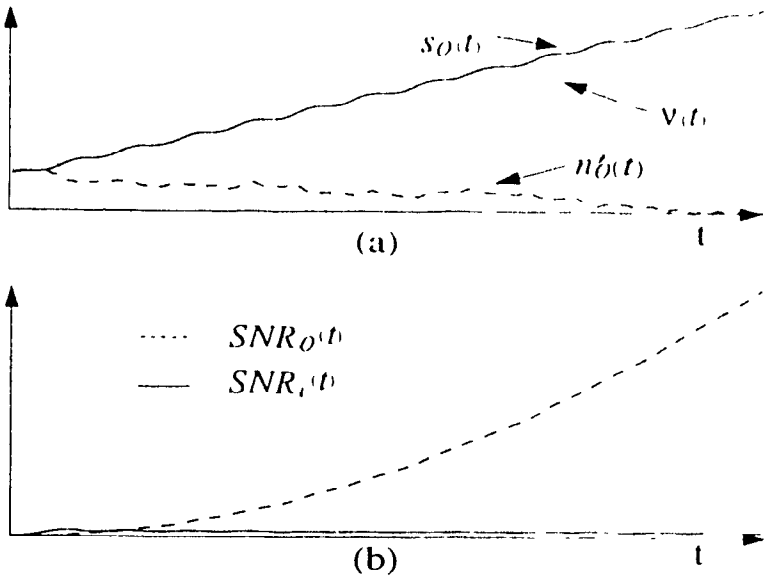


Figure 4.5: (a) $\nu(t)$ and its individual components. (b) the input and output SNRs of IMF as a function of time.

Examining (4.14), we can see that if we now scale $\nu(t)$ by $\phi(t)/c(t)$, then the result represents the desired signal $s(t)$ plus some noise,

$$\hat{s}(t) = [Vc(t) + n'_o(t)]\frac{\phi(t)}{c(t)} = V\phi(t) + n_o(t)$$
$$= s(t) + n_o(t) \qquad (4.16)$$

In fact, the idea of maximization of the SNR_o remains intact. This does not represent or imply optimal estimation of $s(t)$. It does, however, give an intuitive feel for the optimal solution.

To determine the optimal null filter, we scale the output of IMF, $\nu(t)$, by an unknown function $\lambda(t)$ and subtract the result from the input to form the output $y(t)$ as in Figure 1.6,

$$y(t) = x(t) - y'(t) = x(t) - \hat{s}(t) = \hat{n}(t) \tag{4.17}$$

where the scaled output of the IMF, $y'(t)$, represents the estimate of $s(t)$. Replacing (4.2) and (4.14) in (4.17), $y(t)$ can be written as

$$y(t) = s(t) + n(t) - [vc(t) + n_o(t)]\lambda(t). \tag{4.18}$$

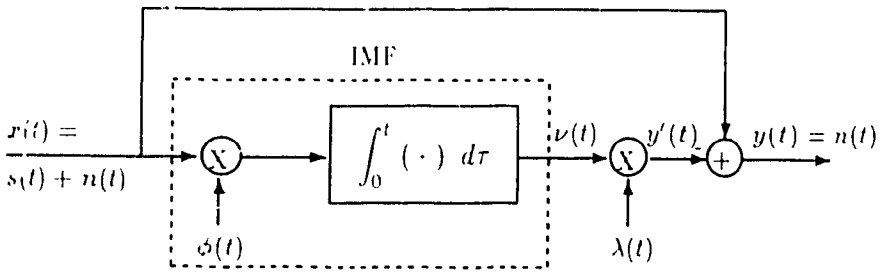


Figure 4.6: Statistically Optimal Null Filter – Coherent Case

It now remains to find a suitable function, $\lambda(t)$ such that $y(t)$ represents the input noise $n(t)$ in the MMSE sense. This is equivalent to estimating $s(t)$ with MMSE.

For ideal null filtering, $y_{ideal}(t) = n(t)$, thus the error in filtering becomes

$$\begin{aligned} e_\lambda(t) &= y_{ideal}(t) - y(t) \\ &= [vc(t) + n_o(t)]\lambda(t) - v\phi(t) \end{aligned} \tag{4.19}$$

Minimizing the mean-square error, $E[e_\lambda^2(t)]$, with respect to the scaling function $\lambda(t)$, yields

$$\lambda_{opt}(t) = \frac{\overline{V^2}\phi(t)}{\overline{V^2}c(t) + N_o} \tag{4.20}$$

where we have used the subscripts 'opt' to denote the optimal scaling function, $E[V^2] = \overline{V^2}$, and N_o is the power of the noise at the input. The optimal post-IMF scaling function,

$\lambda_{opt}(t)$ can be written in terms of the input SNR as,

$$\lambda_{opt}(t) = \frac{\phi(t)}{c(t) + 1/SNR} \tag{4.21}$$

where $SNR = \overline{V^2}/N_o$. Substituting $\lambda_{opt}(t)$ back in (4.19) gives a biased estimate (asymptotically unbiased),

$$E[e_{\lambda_{opt}}(t)] = E[V][c(t)\lambda_{opt}(t) - \phi(t)] \neq 0 \tag{4.22}$$

and the following MMSE

$$E\left[e^2_{\lambda_{opt}}(t)\right] = N_o \frac{\overline{V^2}\phi^2(t)}{\overline{V^2}c(t) + N_o} = N_o\phi(t)\lambda_{opt}(t) \quad \eta_{opt}(t) \tag{4.23}$$

Thus, $\lambda_{opt}(t)$ provides statistically optimal but biased null filtering. We refer to this as the Statistically Optimal Null-Filter (SONF).

Suboptimal Coherent Null Filtering

In order to implement the SONF, the knowledge of the input SNR is required. To circumvent this problem, the following suboptimal approach may be considered. If it is assumed that the input noise is weak (i.e., $N_o \to 0$) then

$$\lambda_{opt}(t) \longrightarrow \lambda'(t) = \frac{\phi(t)}{c(t)} \tag{4.24}$$

where the prime has been used to indicate the suboptimal approach. Notice that $\lambda'(t)$ is exactly the post-IMF scaling function that we obtained by inspection of (4.14) to get an intuitive feel for the approach. By using $\lambda'(t)$, a suboptimal post IMF scaling function, we get unbiased null-filtering with the following MMSE

$$E[e^2_{\lambda'}(t)] = N_o \frac{\phi^2(t)}{c(t)} = \eta'(t). \tag{4.25}$$

In the SONF approach, we have taken an optimal result of the proposed IMF (in the maximum SNR_o sense) and then applied a least-square minimization procedure to find an optimal scaling waveform, $\lambda(t)$, in the hope of achieving an optimal null filter. One may understand it as minimizing the error in the SNR_o estimation of $s(t)$. Comparing the MMSE of the optimal and the suboptimal approaches, we obtain the ratio

$$\gamma(t) = \frac{\eta'(t)}{\eta_{opt}(t)} = 1 + \frac{N_o}{\overline{V^2}c(t)} = 1 + \frac{1}{SNR\, c(t)}. \tag{4.26}$$

The additional information of the signal and noise power in $\lambda_{opt}(t)$, provides an even lower MMSE. This, of course, is at the expense of the knowledge of more input signal information and a biased estimate. From the ratio $\gamma(t)$ it is clear that asymptotically both of the approaches yield equal MMSE. As $t \to \infty$, the second term in (4.26) approaches zero. In so doing, the post IMF scaling function $\lambda_{opt}(t)$ approaches $\lambda'(t)$. The net effect is that *the influence of $\lambda_{opt}(t)$ is mainly felt during the start of the filtering.*

It is well known that the MFs minimize the noise in a very short time interval. From (4.23) and (4.25) this is confirmed as $c(t)$ rapidly increases to a large value forcing a small MSE very quickly.

4.3.2 Noncoherent Null Filter

Up to now we have described a coherent approach of signal suppression where the signal shape is *a priori* known and the main building block is the IMF. Here, we consider a more general scenario, a noncoherent case where the signal shape is unknown.

Consider the case where the signal $s(t)$ can be written as a linear combination of a set of orthogonal basis functions $\{\phi_i(t), i = 1, 2 \ldots N\}$, as in (4.3). We assume that the composing basis functions are *a priori* known while the coefficients v_i's are unknown random variables. In situations where the basis function are unknown, they can always be evaluated or estimated given the second order statistics – the KL expansion for example. In this sense, the problem still remains general. For the moment, we will also assume that the basis functions remain orthogonal for any time interval t, i.e.,

$$\int_0^t \phi_i(\tau)\phi_j(\tau)d\tau = 0 \quad \forall \quad t \in [0, T] \;\; \& \;\; i \neq j. \tag{4.27}$$

This is clearly not possible; however, for the purpose of illustration we will assume it to be so. Later we will show how to deal with this.

Since the filter derived for the coherent case is an LTV filter, the principle of superposition is valid. This filter can, therefore, be implemented as a set of N parallel branches – one to estimate each term in the expansion of $s(t)$ – as shown in Figure 4.7. The post-IMF scaling functions $\lambda_i(t)$ can be chosen to be either optimal or suboptimal as described by (4.20) or (4.24), respectively. Notice that, if we do not assume the orthogonality of the basis functions for any time interval $t \in [0, T]$, then the output of the IMFs at each

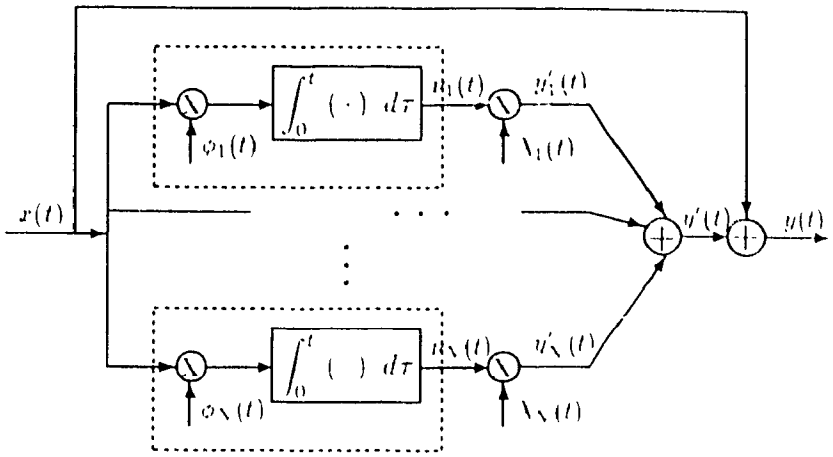


Figure 4.7. Noncoherent Statistically Optimal Null Filter

time instant does not provide a maximum S/N R. If we examine the ith branch of Figure 4.7 independently, (to estimate ith term of $s(t)$) then the interference from the other $N-1$ nonorthogonal terms is not of zero-mean gaussian type. This interference is seen as colored noise, hence, the earlier discussion of the IMF in terms of MF is not valid. That is, the output of the IMF in the ith branch can be written as

$$v_i(t) = \sum_{j=1}^{N} \int_0^t \phi_i(\tau)\phi_j(\tau)d\tau \int_0^t u(\tau)\phi_i(\tau)d\tau \qquad (4.28)$$

It is clear that if we do not use the above orthogonality condition then

$$v_i(t) \neq \int_0^t \phi_i^2(\tau)d\tau + \int_0^t n(\tau)\phi_i(\tau)d\tau. \qquad (4.29)$$

hence, $y_i'(t)$ cannot be the estimate of $s_i(t)$, the ith component of $s(t)$.

Interestingly, Van Trees's [44] solution of optimal estimation of $s(t)$ (summarized earlier in the preliminary section) is based on the eigenvalues and the eigenvectors of the kernel or covariance function of $s(t)$. In this approach, a fixed interval of integration is used (observe the complete signal then estimate), whereas in the proposed approach it is varying.

For the case where the post-IMF scaling functions are optimal, $\lambda_i(t) \quad \lambda_{i,opt}(t)$ as

in (4.20), we can write the estimation error as

$$\epsilon_{\lambda,opt}(t) = \sum_{i=1}^{N} \left[\frac{n_{oi}(t)\overline{v_i^2}\phi_i(t)}{\overline{v_i^2}c_i(t) + N_o} - N_o \frac{v_i\phi_i(t)}{\overline{v_i^2}c_i(t) + N_o} \right]. \tag{4.30}$$

Asymptotically, the suppression of $s(t)$ is unbiased with a MMSE given by,

$$\begin{aligned} E\left[\epsilon^2_{\lambda,opt}(t)\right] &= N_o \sum_{i=1}^{\Lambda} \phi_i(t)\lambda_{i,opt}(t) \\ &= N_o \sum_{i=1}^{N} \frac{\overline{v_i^2}}{\overline{v_i^2}c(t) + N_o} \phi_i^2(t) = \eta_{opt}(t) \end{aligned} \tag{4.31}$$

which is also zero asymptotically. The approach of Van Trees [11] yields biased estimation with MMSE

$$\epsilon(t) = N_o \sum_{i=1}^{N} \frac{\mu_i}{\mu_i + N_o} \phi_i^2(t) \qquad 0 \le t \le T \tag{4.32}$$

In contrast to our approach, the MMSE in this case, however, does not approach zero for $t \to \infty$.

Suboptimal Noncoherent Null Filtering

If the suboptimal $\lambda(t)$ is used in Figure 4.7 i.e., $\lambda(t) = \lambda'(t)$, we achieve an unbiased estimate with a MMSE that is initially poorer than $\eta_{opt}(t)$, but still remains consistent:

$$\lim_{t\to\infty} E\left[\epsilon_N^2(t)\right] = \lim_{t\to\infty} N_o \sum_{i=1}^{N} \frac{\phi_i^2(t)}{c_i(t)} = 0. \tag{4.33}$$

It should be noted that in using $\lambda_{i,opt}(t)$ in Figure 4.7, the Noncoherent-SONF (NC-SONF) is only locally optimal. That is, each component of the signal $s(t)$ is estimated independently with the assumption that no other terms exist. This does not provide the global optimality of the filter in Figure 4.7. In the next section, we verify this hypothesis by constructing a globally optimal SONF.

4.3.3 Orthogonalization

In the noncoherent SONF approach, it was assumed that a linear expansion of the signal $s(t)$ using **orthogonal basis functions** is available. This assumption was made even more strict in that it was required for the basis set $\{\phi_i(t), i = 1, 2 \ldots N\}$ to be orthogonal

for any arbitrary time interval $t \geq T$, (see (4.27)). This assumption cannot be satisfied in practice. One method of overcoming this difficulty is to orthogonalize the functions $\phi_i(t)$'s for each new instant of time interval, t, as t is increasing. We propose to use a *sliding* Gram-Schmidt (GS) orthogonalization procedure described in the preliminary section. By applying the sliding GS procedure, the expansion of the signal $s(t)$, in terms of the set $\{\phi_i(t), i = 1, 2 \ldots N\}$, is transformed into another representation with a new basis set $\{v_i(t), i = 1, 2 \ldots N\}$ that are orthogonal over any interval $t \in [0, T]$. With this orthogonalization, the constant coefficient expansion of $s(t)$ in (4.3) is transformed into an expansion where the coefficients are now time-varying as in (4.13).

4.4 Globally Optimal SONF

In using the SONF, a linear expansion of the signal to be suppressed/enhanced is required. The basis functions in the expansion need to be orthogonal for any time interval $t \in [0, T]$. Since this is clearly not possible, as a possible solution in the last section, we have proposed a sliding GS orthogonalization of the basis functions at each observation interval as the observation interval increases. This approach is effective, however, the computations involved are greatly increased. To deal with the increased number of computations, we try to find in this section, post-IMF scaling functions, $\lambda_i(t)$'s wherein the orthogonalization procedure is unnecessary. We remove the assumption of orthogonality of the basis functions. This will cause errors in the expected IMF output, however, these errors can be minimized by the optimization of $\lambda(t)$'s. Moreover, in so doing, we will obtain a globally optimal solution of $\lambda_i(t)$'s. That is, all the $\lambda_i(t)$ are calculated simultaneously.

To find the $\lambda_{i,opt}(t)$ that are globally optimal, we reconsider the problem in vector notation. Referring to Figure 4.7, we define the following:

$$\begin{aligned} \lambda(t) &= [\lambda_1(t)\lambda_2(t)\cdots\lambda_N(t)]^T \\ \nu(t) &= [\nu_1(t)\nu_2(t)\cdots\ \nu_N(t)]^T \\ \phi(t) &= [\phi_1(t)\phi_2(t)\cdots\phi_N(t)]^T \\ V &= [v_1 v_2 \cdots v_N]^T \end{aligned} \tag{4.34}$$

where $\lambda(t)$, $\nu(t)$, $\phi(t)$ and V are the post-IMF scaling functions, output of the IMF, the

set of known basis functions and the amplitude of the each term of the desired signal, respectively. By using (4.31), the input signal can be written as

$$x(t) = V^T \phi(t) + n(t) \tag{4.35}$$

and the output signal as

$$y(t) = V^T \phi(t) + n(t) - \lambda^T(t)\nu(t) \tag{4.36}$$

The error in suppressing $s(t)$ becomes

$$\epsilon(t) = n(t) - n(t) = V^T \phi(t) - \lambda^T(t)\nu(t) \tag{4.37}$$

and the MSE as a function of $\lambda(t)$ can be written as,

$$\begin{aligned} E\left[\epsilon^2_\lambda(t)\right] &= \eta_\lambda(t) \\ &= \lambda^T(t)Q(t)\lambda(t) - \lambda^T(t)\rho(t) - \rho^T(t)\lambda(t) + V^T\phi(t)\phi^T(t)V \end{aligned} \tag{4.38}$$

where

$$Q(t) = E\left[\nu(t)\nu^T(t)\right] \tag{4.39}$$

and

$$\rho(t) = E\left[\nu(t)\phi^T(t)V\right]. \tag{4.40}$$

Minimizing (4.38) w.r.t. $\lambda(t)$, i.e.,

$$\left.\frac{\partial \eta_\lambda(t)}{\partial \lambda(t)}\right|_{\lambda(t)=\lambda^g_{opt}(t)} = 0 = Q(t)\lambda(t) - \rho(t), \tag{4.41}$$

yields the globally optimal solution for $\lambda(t)$,

$$\lambda^g_{opt}(t) = [\mathcal{V}D(t) + N_o]^{-1}\,[\mathcal{V}\phi(t)] \tag{4.42}$$

where we have used the superscript 'g' to denote globally optimal and

$$D(t) = \int_0^t \phi(\tau)\phi^T(\tau)d\tau \tag{4.43}$$

and

$$\mathcal{V} = E\left[VV^T\right]. \tag{4.44}$$

Consider the case for $N = 2$, i.e., a two-term representation of $s(t)$,

$$s(t) = [v_1 v_2][\phi_1(t)\phi_2(t)]^T = V\phi^T(t). \tag{4.45}$$

Equation (4.42) can be evaluated to yield

$$\lambda^g_{1,opt}(t) = \frac{(ab-c^2)\left[\beta_2(t)\phi_1(t) - \beta_{12}(t)\phi_2(t) + aN_o\phi_1(t) + c\phi_2(t)\right]}{(ab-c^2)\left[\beta_1(t)\beta_2(t) - \beta^2_{12}(t)\right] + a\beta_1(t)N_o + b\beta_2(t)N_o + 2c\beta_{12}(t)N_o + N_o^2} \tag{4.46}$$

and

$$\lambda^g_{2,opt}(t) = \frac{(ab-c^2)\left[\beta_1(t)\phi_2(t) - \beta_{12}(t)\phi_1(t) + N_o c\phi_1(t) + bN_o\phi_2(t)\right]}{(ab-c^2)\left[\beta_1(t)\beta_2(t) - \beta^2_{12}(t)\right] + a\beta_1(t)N_o + b\beta_2(t)N_o + 2c\beta_{12}(t)N_o + N_o^2} \tag{4.47}$$

where

$$\begin{aligned} \beta_1(t) &= \int_0^t \phi_1^2(\tau)d\tau \\ \beta_2(t) &= \int_0^t \phi_2^2(\tau)d\tau \\ \beta_{12}(t) &= \int_0^t \phi_1(\tau)\phi_2(\tau)d\tau \\ a &= E[v_1^2] \\ b &= E[v_2^2] \\ c &= E[v_1 v_2] \end{aligned} \tag{4.48}$$

In the numerator of expressions for $\lambda^l_{1,opt}(t)$ and $\lambda^l_{2,opt}(t)$, some form of correlation removal can be seen from the $\beta_{12}(t)$ terms. We have some form of indirect orthogonalization. If the basis functions are orthogonal with the definition of (4.27), then $\lambda^l_{1,opt}(t)$ and $\lambda^g_{2,opt}(t)$ reduce to

$$\lambda^g_{1,opt}(t) = \frac{(ab-c^2)\left(\beta_2(t)\phi_1(t) + aN_o\phi_1(t) + cN_o\phi_2(t)\right)}{(ab-c^2)\beta_1(t)\beta_2(t) + a\beta_1(t)N_o + bN_o\beta_2(t) + N_o^2} \tag{4.49}$$

$$\lambda^g_{2,opt}(t) = \frac{(ab-c^2)\left(\beta_1(t)\phi_2(t) + bN_o\phi_2(t) + cN_o\phi_1(t)\right)}{(ab-c^2)\beta_1(t)\beta_2(t) + a\beta_1(t)N_o + bN_o\beta_2(t) + N_o^2} \tag{4.50}$$

which are clearly not the same as $\lambda_{i,opt}(t)$ of (4.20). Notice the difference between the post-IMF scaling function of the locally and globally optimal filters, thus verifying the hypothesis of the difference in globally and locally optimal SONFs.

4.5 Discrete-Time NC-SONF

To implement the exact expressions of the globally optimal post IMF scaling function, $\lambda^g_{i,opt}(t)$ of (4.42), requires the knowledge of the noise and individual signal component

power, as seen for the case of $N = 2$ in equations (4.46) and (4.47). This is generally not possible in practice. To avoid the calculation of $\lambda^{\gamma}_{i,opt}(t)$'s, we will now derive a discrete-time recursive implementation of the SONFs that is globally optimal.

4.5.1 Non-Coherent Case

Until now we have considered the SONF in continuous-time. Here we derive a recursive form of the NC-SONF for the case of unknown signal shape. To do so, we switch from continuous time to discrete-time formulation. The changeover is straight forward, we simply replace the integration in the IMF by a summation and use the discrete independent time variable n. Figure 4.8 outlines the discrete version of the NC-SONF. Equation (4.31) can be written using discrete independent variable n as,

$$\begin{aligned} \lambda(n) &= [\lambda_1(n)\lambda_2(n)\cdots\lambda_N(n)]^T \\ \nu(n) &= [\nu_1(n)\nu_2(n)\cdots\nu_N(n)]^T \\ \phi(n) &= [\phi_1(n)\phi_2(n)\cdots\phi_N(n)]^T \\ \Psi &= [e_1 e_2 \cdots e_N]^T \end{aligned} \tag{4.51}$$

The input mixture, $x(n)$, can now be written as,

$$x(n) = \Psi^T \phi(n) + n(n). \tag{4.52}$$

The output of the IMF is,

$$\begin{aligned} \nu(n) &= \sum_{m}^{N} x(m)\phi(m) \\ &= \sum_{m}^{n-1} x(m)\phi(m) + x(n)\phi(n) \\ &= \nu(n-1) + x(n)\phi(n) \end{aligned} \tag{4.53}$$

and the output $y'(n)$ (estimate of $s(n)$) can be written as,

$$y'(n) = \lambda^T(n)\nu(n) = \nu^T(t)\lambda(n). \tag{4.54}$$

We evaluate the post-IMF scaling function $\lambda(n)$, in the LMS sense as for the discrete-time

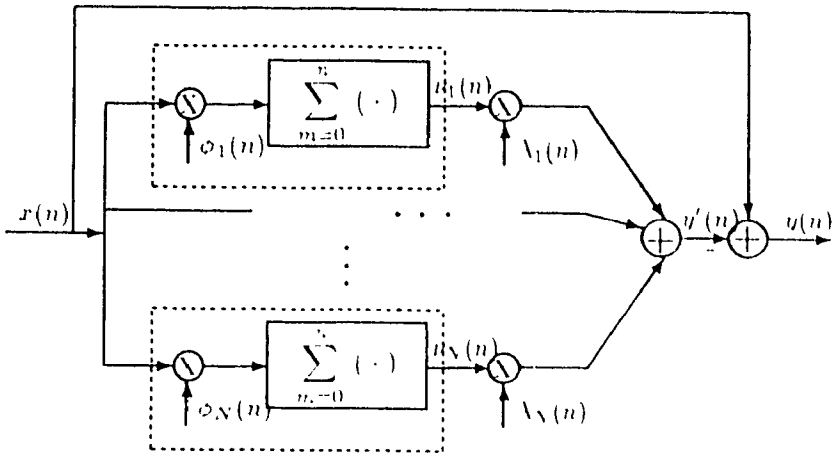


Figure 1.8: Discrete version of the Non Coherent SONF

case. Referring to Figure 1.8, we formulate the error function

$$\begin{aligned} e(n) &= y(n) - u(n) \quad s(n) \quad s(n) \\ &= \lambda^T(n)\nu(n) \quad V^T\varphi(n) \end{aligned} \tag{1.55}$$

and the MSE can be written as,

$$\begin{aligned} \eta(n) &= E[e^2(n)] = E[e(n)e^T(n)] \\ &= \lambda^T(n)Q(n)\lambda(n) - \lambda^T(n)\varphi(n) \quad \varphi^T(n)\lambda(n) + V^T\varphi(n)\varphi^T(n)V \end{aligned} \tag{1.56}$$

where

$$Q(n) = E\left[\nu^T(n)\nu(n)\right] \tag{1.57}$$

$$\rho(n) = E\left[\nu(n)\phi^T(n)V\right] \tag{1.58}$$

By using the following facts [69],

$$\begin{aligned} \frac{\partial}{\partial\lambda(n)}\left[\lambda^T(n)Q(n)\lambda(n)\right] &= 2Q(n)\lambda(n) \\ \frac{\partial}{\partial\lambda(n)}\left[\rho^T(n)\lambda(n)\right] &= 0 \\ \frac{\partial}{\partial\lambda(n)}\left[\lambda^T(n)\rho(n)\right] &= 2\rho(n) \end{aligned}$$

we can minimize (4.56) w.r.t. $\lambda(n)$ as,

$$\left.\frac{\partial \eta(n)}{\partial \lambda(n)}\right|_{\lambda(n)=\lambda^{I}_{opt}(n)} = 2Q(n)\lambda^{I}_{opt}(n) - 2\rho(n) = 0. \tag{4.59}$$

This yields a system of linear equations that can be solved for $\lambda^{I}_{opt}(n)$, i.e.,

$$Q(n)\lambda^{I}_{opt}(n) = \rho(n). \tag{4.60}$$

The matrix $Q(n)$ and $\rho(n)$ is evaluated to be

$$Q(n) = D(n)\mathcal{V}D(n) + \mathcal{N}_o D(n) \tag{4.61}$$

$$\rho(n) = D(n)\mathcal{V}\phi(n) \tag{4.62}$$

where

$$\begin{aligned} D(n) &= \sum_{m=1}^{n} \phi(m)\phi^{I}(m) \\ &= D(n-1) + \phi(n)\phi^{I}(n) \end{aligned} \tag{4.63}$$

Rewriting (4.60) in terms of $Q(n)$ and $\rho(n)$ and with the assumption that $D(n)$ is positive definite, we can solve for $\lambda^{I}_{opt}(n)$,

$$\lambda^{I}_{opt}(n) = [\mathcal{V}D(n) + \mathcal{N}_o I]^{-1}\mathcal{V}\phi(n) \tag{4.64}$$

We have thus derived globally optimal post-IMF scaling functions, $\lambda^{I}_{pt}(n)$. Notice that, the original restriction of the orthogonality of the basis function for any time $t \in [0, T]$ is not used in this development. In performing a global optimization of the post-IMF scaling functions, i.e. $\lambda^{I}_{opt}(n)$, an implicit orthogonalization has been done. This phenomenon was described in the previous section in the continuous-time case for $N = 2$, which is equally valid for the discrete-time case. We have, thus, eliminated the need for this strict orthogonalization requirement. It now remains to show that we can implement $\lambda^{I}_{opt}(n)$ recursively to eliminate the need for the knowledge of $\mathcal{V}$ and $\mathcal{N}_o$ in (4.64).

4.5.2 Recursive Implementation

Let

$$P(n) = R(n)^{-1}\mathcal{V} \tag{4.65}$$

with

$$R(n) = \gamma D(n) + \lambda, \tag{4.66}$$

Now, by using (4.65) and (4.66), (4.64) can be written as

$$\lambda^T(n) = P(n)\phi(n) \tag{4.67}$$

and $P(n)$ can be evaluated recursively. Substituting (4.63) in (4.66), we obtain

$$\begin{aligned} R(n) &= \gamma D(n-1) + \lambda + \gamma\phi(n)\phi^T(n) \\ &= R(n-1) + \gamma\phi(n)\phi^T(n) \end{aligned} \tag{4.68}$$

Using (4.65) and (4.68), we rewrite $P(n)$ in terms of its past value as

$$P(n) = \left[P(n-1)^{-1} + \phi(n)\phi^T(n)\right]^{-1} \tag{4.69}$$

A recursive update formula for the gain matrix $P(n)$ can be found with the aid of the matrix inversion lemma [69]

$$[A + BC^{-1}]^{-1} = A^{-1} - A^{-1}B[DA^{-1}B + C^{-1}]^{-1}DA^{-1} \tag{4.70}$$

where A and C are positive definite matrices. Letting

$$\begin{aligned} A &= P(n-1)^{-1} \\ B &= \phi(n) \\ C &= 1 \\ D &= \phi^T(n) \end{aligned}$$

in (4.70), P(n) can be calculated to be

$$P(n) = P(n-1) - \frac{P(n-1)\phi(n)\phi^T(n)P(n-1)}{1 + \phi^T(n)P(n-1)\phi(n)} \tag{4.71}$$

The complete recursive algorithm for implementing the SONF can be written as

$$\nu(n) = \nu(n-1) + x(n)\phi(n) \tag{4.72}$$

$$P(n) = P(n-1) - \frac{P(n-1)\phi(n)\phi^T(n)P(n-1)}{1 + \phi^T(n)P(n-1)\phi(n)} \tag{4.73}$$

$$\lambda(n) = P(n)\phi(n) \tag{4.74}$$

$$y(n) = x(n) - \nu^T(n)\lambda(n). \tag{4.75}$$

The gain matrix $P(n)$ is initially chosen to be positive definite. As a rule of thumb one may choose $P(0) = SNR\, I$, where I is the identity matrix of order N and $\epsilon(0) = 0$.

4.6 Adaptive SONF

Transient signals are often modeled as a combination of damped sinusoids. Thus, to generate the basis functions for implementing the SONF, knowledge of the damping coefficient is assumed to be *a priori* known. However, this is generally not the case as in the modeling of evoked potentials [70]. Objective in this section is to present a method of adaptively estimating the damping coefficients of damped sinusoids using the newly developed SONF.

For simplicity, a single damped sinusoid case is considered. Let

$$s(n) = A e^{-\beta n} \cos(\omega_o n + \theta) = c_1 \phi_1(n) + c_2 \phi_2(n) \tag{4.76}$$

be the damped sinusoid of interest, where $\phi_1(n) = e^{-\beta n}\cos(\omega_o n)$ and $\phi_2(n) = e^{-\beta n}\sin(\omega_o n)$ are the basis functions, and c_1 and c_2 are unknown coefficients in the expansion. It is assumed that the sinusoid frequency ω_o is *a priori* known and A, θ, β are the unknown amplitude, initial phase and damping coefficient. The received or measured signal is corrupted by zero-mean AWGN,

$$x(n) = s(n) + w(n) \tag{4.77}$$

Since the basis functions are not completely known, we aim to use the SONF in an adaptive form to extract the damped sinusoid from the received signal. The basis functions $\phi_i(n)$'s can be formulated from the on-line estimate of the damping coefficient as,

$$\hat{\phi}_1(n) = e^{-\hat{\beta}(n)n}\cos(\omega_o n) \tag{4.78}$$

$$\hat{\phi}_2(n) = e^{-\hat{\beta}(n)n}\sin(\omega_o n) \tag{4.79}$$

where $\hat{\beta}(n)$ is the current estimate of the damping coefficient. Complete adaptive SONF structure is described by the block diagram of Figure 4.9. To estimate the damping coefficient, we can use the stochastic gradient method. With this method, the update equation for the damping coefficient can be written as,

$$\hat{\beta}(n+1) = \hat{\beta}(n) + \gamma \Psi(n) \epsilon(n) \tag{4.80}$$

where the hat indicates an estimated quantity, γ is the rate of convergence (step-size), $\Psi(n)$ the gradient and $e(n)$ is the error in the current estimate of $s(n)$.

Since $x(n)$ the desired signal corrupted by noise, referring to Figure 4.9, we can write the following expression for the fitting error

$$e(n) = x(n) - s(n) = x(n) - \lambda^T(n)\nu(n) \tag{4.81}$$

Note that both $\lambda(n)$ and $\nu(n)$ are dependent on β. The negative gradient of $e(n)$ w.r.t. β can be found as follows. Defining the negative gradient of $e(n)$ as,

$$\Psi(n) = -\frac{\partial e(n)}{\partial \beta} = -\frac{\partial [x(n) - \lambda^T(n)\nu(n)]}{\partial \beta}$$
$$= \frac{\partial \lambda^T(n)\nu(n)}{\partial \beta} - \frac{\partial \lambda^T(n)}{\partial \beta}\nu(n) + \lambda^T(n)\frac{\partial \nu(n)}{\partial \beta} \tag{4.82}$$

which can be evaluated to be (see Appendix)

$$\Psi(n) = \sum_{i=1}^{2} \frac{\phi_i(n)}{A_i(n)} \left[n\nu_i(n) - 2\frac{\nu_i(n)}{A_i(n)} B_i(n) + C_i(n) \right] \tag{4.83}$$

where $\nu_i(n)$ is as defined earlier, see (4.53), and

$$\begin{aligned} A_i(n) &= A_i(n-1) + \phi_i^2(n) \\ B_i(n) &= B_i(n-1) + n\phi_i^2(n) \\ C_i(n) &= C_i(n-1) + nx(n)\phi_i(n) \end{aligned} \tag{4.84}$$

The complete adaptive SONF to estimate the damping coefficient can be written as,

$$y(n) = \lambda^T(n)\nu(n) \tag{4.85}$$

$$\epsilon(n) = x(n) - y(n) \tag{4.86}$$

$$\Psi(n) = \sum_{i=1}^{2} \frac{\phi_i(n)}{A_i(n)} \left[n\nu_i(n) - 2\frac{\nu_i(n)}{A_i(n)} B_i(n) + C_i(n) \right] \tag{4.87}$$

$$\hat{\beta}(n+1) = \hat{\beta}(n) + \gamma\Psi(n)e(n). \tag{4.88}$$

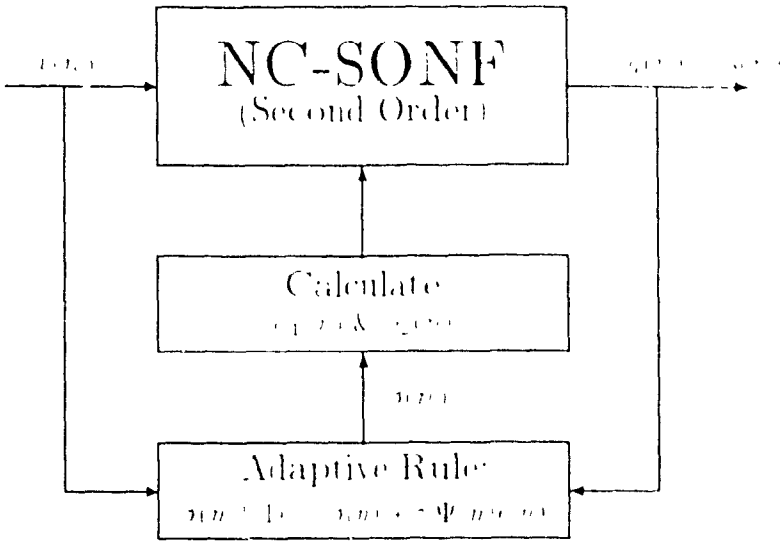


Figure 4.9 Adaptive SONF with a stochastic gradient update algorithm

4.7 Analysis: Kalman Equivalence of the SONF

In the previous sections, we presented a new method of suppression/enhancement of narrowband (NB) signals. The new method and its variants were based on the basic building block of the IMF. Since the new filters are time varying and are optimized under the MSE criterion, it is anticipated that they may be closely related to the well known Kalman Filter, first presented in [71]. In this section, we will first briefly outline the basic Kalman Filter and then show the relationship of the SONF to the Kalman Filter. We will analytically show this relationship for the coherent SONF (first order) case. For the higher order or the NC-SONF, the mathematics involved become intractable, as such we will only formulate the mathematical setting and show via simulation its Kalman equivalence.

4.7.1 Kalman Filter

Kalman filter theory provides an alternative way of formulating the least square filtering problem. It provides the solution to a class of recursive minimum mean-square estimation problems [71]. Two main features of the Kalman filter are its 1) vector modeling of random processes under consideration (state-space approach) and 2) the recursive processing of observed or measured noisy data [72, 73]. We now briefly summarize the Kalman Filter.

The notation used here is consistent with that used in [71].

Problem Description

The scalar process $y(t)$ under analysis can be thought of as a linear combination of states of some dynamical system. The signal model can be described by the following dynamical system

$$\dot{x}(t) = F(t)x(t) + G(t)u(t) \tag{4.89}$$

$$y(t) = H^T(t)x(t) + v(t) \tag{4.90}$$

where $F(t)$, $G(t)$ and $H(t)$ are rectangular matrices with dimensions n x n, n x m and n x p respectively and $u(t)$ is the input to the dynamical system. The observed signal $z(t)$ is an additive mixture of the process under analysis ($y(t) = H^T(t)x(t)$) and noise $v(t)$. The processes $u(t)$ and $v(t)$ are uncorrelated zero mean white noise type.

Unlike the Wiener filter, the Kalman filter does not estimate the process $x(t)$ directly. Instead it estimates the state vector $x(t)$ from the observed signal $z(t)$. The following succinctly states the Kalman problem:

Given the observed values of $z(\tau)$ in the interval $t_0 \le \tau \le t$, the task is to find an estimate $\hat{x}(t)$ of $x(t)$ of the form

$$\hat{x}(t) = \int_{t_0}^{t} \Lambda(t,\tau)z(\tau)d\tau \tag{4.91}$$

(where Λ is an $n \times p$ matrix whose elements are continuously differentiable in both arguments) with the property that the mean-squared error is minimized:

$$E\left[(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T\right] = \text{minimum}. \tag{4.92}$$

Moreover, $\hat{x}(t)$ is to be an on-line estimate.

Kalman Solution

The now well-known Kalman solution to optimal filtering is shown in Figure 4.10. The optimal estimate $\hat{x}(t)$ is generated by a linear dynamical system of the form

$$\dot{\hat{x}}(t) = F(t)\hat{x}(t) + K(t)\tilde{z}(t) \tag{4.93}$$

$$\tilde{z}(t) = z(t) - H^T(t)\hat{x}(t). \tag{4.94}$$

The initial state $x(t_0)$ of (4.93) is a zero vector. This optimal filter is a feedforward system. It is obtained by taking the message process model (4.89) (omitting the input $w(t)$) and feeding forward the error in the estimate of $y(t)$ with a gain $K(t)$ (see Figure 4.10). Thus, specifying the Kalman Filter is equivalent to computing the feedforward gain $K(t)$. Defining the symmetric matrix $P(t_0) = E\{[x(t_0) - \hat{x}(t_0)][x(t_0) - \hat{x}(t_0)]^T\}$, the optimal

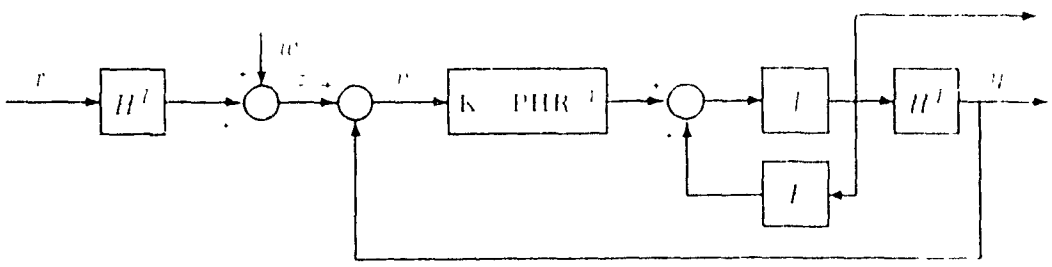


Figure 4.10 Kalman Filter

gain (Kalman gain) has been shown to be [71]

$$K(t) = P(t)H^T(t)R^{-1}(t) \tag{4.95}$$

where $R(t) = E[v(t)v^T(t)]$ is the observed noise covariance matrix. The only remaining unknown is the matrix $P(t)$, the covariance matrix in the estimate of $x(t)$. It can be found as the solution to the nonlinear matrix differential equation of the Riccati type

$$\dot{P}(t) = F(t)P(t) + P(t)F^T(t) - P(t)H^T(t)R^{-1}(t)H(t)P(t) + G(t)Q(t)G^T(t) \tag{4.96}$$

where $Q(t) = E[w(t)w^T(t)]$. Together (4.93), (4.94), (4.95) and (4.96) comprise the Kalman Filter. Furthermore, the performance of the Kalman estimator (4.93) is given by $P(t)$.

4.7.2 Kalman Filter Equivalence of SONF

In order for this part to be complete on its own, we have chosen to repeat some of the equations presented earlier. For the first order case, the notation used here is consistent with that used for the description of the coherent SONFs throughout this chapter. For the noncoherent case slight modifications are required. Care must be exercised not to confuse with the notation used in Section 4.7.1.

First Order

The estimation counterpart of the first order SONF or the coherent SONF is shown in Figure 4.11 and its output can be written as

$$y(t) = \lambda(t)\phi(t)$$

$$\lambda(t) \int_0^t \phi(\tau) x(\tau) d\tau \tag{4.97}$$

where $x(t)$ is the observed signal i.e. $x(t) = s(t) + n(t)$ and $y(t)$ is the estimate of the desired signal i.e. $y(t) = s(t)$. The performance or the MMSE in the estimate of $s(t)$ is found to be (see section 4.3)

$$\dot{y}(t) = \dot{\lambda}(t) \int_0^t \phi(\tau) x(\tau) d\tau + \lambda(t)\phi(t)x(t) \tag{4.98}$$

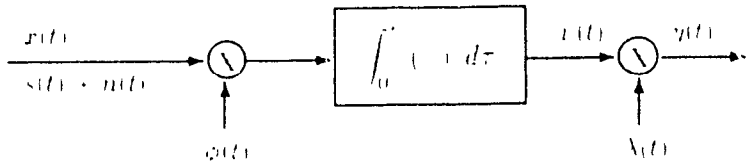


Figure 4.11 Signal estimation counterpart of the coherent SONF

In the following, we show how (4.97) and (4.98) comprise the equivalence of the Kalman Filter. We will first derive a differential equation of which $\hat{y}(t)$ is a solution (like (4.93) generates the state estimates in the Kalman Filter) and then show how the performance of the SONF can be described by the solution of a nonlinear differential equation (like (4.96) does for the Kalman filter).

Differentiating (4.97), we obtain

$$\dot{y}(t) = \dot{\lambda}(t) \int_0^t x(\tau)\phi(\tau)d\tau + \lambda(t)\phi(t)x(t). \tag{4.99}$$

To form a first order differential equation of the type (4.93), we proceed as follows. Multiplying (4.99) by $\lambda(t)$ gives,

$$\lambda(t)\dot{y}(t) = \lambda(t)\dot{\lambda}(t) \int_0^t x(\tau)\phi(\tau)d\tau + \lambda^2(t)\phi(t)x(t) \tag{4.100}$$

and multiplying (4.97) by N_o gives

$$[illegible] \tag{4.100}$$

By subtracting (4.101) from (4.100) and solving for [illegible], we obtain the differential equation of the type (4.93)

$$[illegible] \tag{4.102}$$

In the case of optimal filtering (see Section [illegible]), the [illegible] IMF [illegible] can be written as

$$[illegible] \tag{4.103}$$

The derivative of $\lambda(t)$ can be evaluated to be

$$[illegible] \tag{4.104}$$

for [illegible] $\neq 0$. Letting [illegible], (4.104) can be written as

$$[illegible] \tag{4.105}$$

Consequently

$$[illegible] \tag{4.106}$$

Using $\phi(t)\lambda(t) = \eta(t)/N_o$ from (4.98)

$$\eta(t) = \left[a(t) - \frac{\eta(t)}{N_o}\right]\eta(t) + \frac{\eta(t)}{N_o} r(t) \tag{4.107}$$

We can rewrite the above in the form of a Kalman Filter,

$$\eta(t) = a(t)\eta(t) + K(t)[r(t) - \eta(t)], \tag{4.108}$$

where $K(t) = \eta(t)/N_o$. Table 4.1 compares or relates the notation of (4.108) to the Kalman Filter of (4.93) and (4.95). Observe that (4.108) is identical to the Kalman State Estimator of (4.93) with the appropriate notational changes, the difference being that in our case, we have a scalar equation rather than a matrix one.

It now remains to show that $\eta(t)$ is the solution of a nonlinear variance differential equation of the type in (4.96). To do so, we differentiate (4.98)

$$\dot{\eta}(t) = N_o\dot{\phi}(t)\lambda(t) + N_o\phi(t)\dot{\lambda}(t) \tag{4.109}$$

Kalman Filter		SONF	
$x(t)$	State Vector	$q(t)$	State vector
$q(t)$	Output	$q(t)$	Output signal
$z(t)$	Observed signal	$z(t)$	Observed signal
$F(t)$		$a(t)$	
$G(t)$		0	
$H(t)$		1	
$P(t)$		$\eta(t)$	
$R(t)$		N	
$K(t)$	$P(t)H^T(t)R^{-1}(t)$	$K(t)$	$\eta(t)$ N

Table 4.1 Notational differences in the Kalman and the coherent SONF

Substituting (4.105) for N (t) yield

$$\eta(t) = N_0 a(t)\lambda(t) + N_0 a(t)\, a(t)\lambda(t) - a(t)\lambda^2(t) \tag{4.110}$$

By using the fact $\lambda(t) = \eta(t) - N_0 a(t)$ (4.110) can be written as

$$\eta(t) = 2a(t)\eta(t) - \frac{\eta^2(t)}{N_0} \tag{4.111}$$

Comparing (4.111) with (4.96) with the appropriate notational differences in Table 4.1, we see that they are identical. Equation (4.111) is a Riccati type of nonlinear differential equation, the solution of which describes the performance of the estimator. Figure 4.12 shows the Kalman Filter like structure of the SONF represented by (4.108) and(4.111)

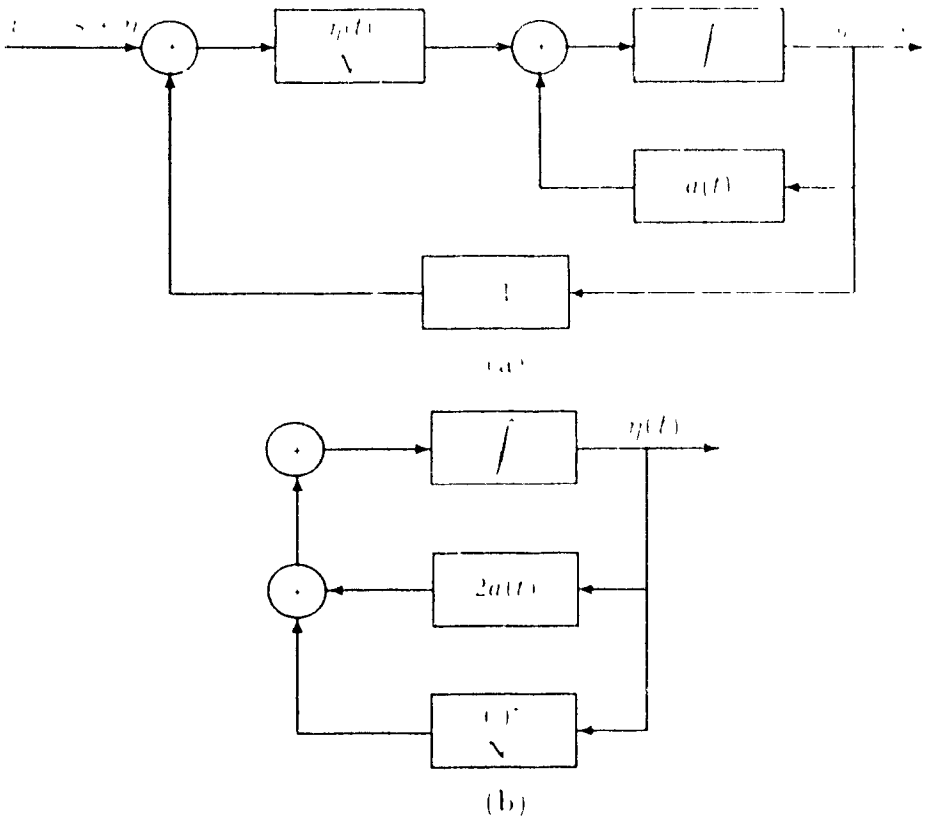


Figure 4.12: Kalman Equivalence of First Order SONEs (a) Dynamical system to estimate the signal; (b) MSE generator.

Generalization to an Nth order case

We now consider the Kalman equivalence of the general Nth order SONF i.e., NC-SONF. Figure 4.13 shows the signal estimation counterpart of the NC-SONF procedure. For the purpose of this section, we will make slight notational changes in the formulation of the NC-SONF. Here, $z(t)$ represents the estimated signal $s(t)$ and $\hat{y}_i(t)$ can be considered the elements of the estimated state vector $y(t)$.

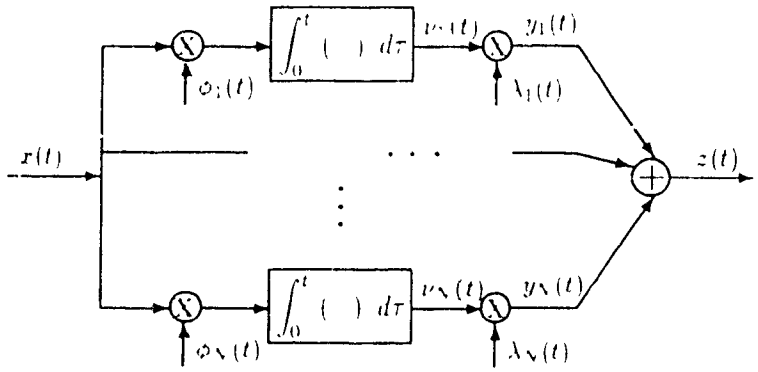


Figure 4.13 Signal estimation counterpart of the NC-SONF.

With reference to Figure 4.13, we reformulate the NC-SONF in a matrix format as follows. Let,

$$\phi(t) = [\phi_1\ \phi_2\ \cdots\ \phi_N]^T \tag{4.112}$$

$$\hat{y}(t) = [\hat{y}_1(t)\ \hat{y}_2(t)\ \cdots\ \hat{y}_N(t)]^T \tag{4.113}$$

$$\nu(t) = [\nu_1(t)\nu_2(t)\cdots\nu_N(t)]^T \tag{4.114}$$

$$\Lambda(t) = diag[\lambda_1\lambda_2\cdots\lambda_N] \tag{4.115}$$

$$C = [1\ 1\ \cdots\ 1]^T \tag{4.116}$$

where all vectors are n x 1 and diag(·) refers to a diagonal matrix of dimension n x n with entries described by the argument of $diag(\cdot)$. Using the above definitions, we can write

the output of our filter in a matrix format as

$$\hat{y}(t) \qquad \Lambda(t)\hat{y}(t) \tag{4.117}$$

$$\hat{z}(t) \qquad C^T\hat{y}(t) \tag{4.118}$$

where we have used the hat to denote estimated quantities. The actual desired signal $z(t) = C^T y(t)$ is assumed to be generated by a linear dynamical system of the type

$$\dot{y}(t) \qquad \Lambda(t)y(t) \tag{4.119}$$

$$z(t) \qquad C^T y(t) \tag{4.120}$$

where $\Lambda(t)$ is an n x n system matrix that is unknown but assumed to exist.

To show the Kalman equivalence of (4.117) and (4.118), we proceed as with the first order case. Differentiating (4.117) we obtain

$$\dot{y}(t) = \dot{\Lambda}(t)\hat{y}(t) + \Lambda(t)\dot{\hat{y}}(t) \tag{4.121}$$

Multiplying (4.121) by $\Lambda(t)$ gives,

$$\Lambda(t)\dot{y}(t) = \Lambda(t)\dot{\Lambda}(t)\hat{y}(t) + \Lambda^2(t)\dot{\hat{y}}(t) \tag{4.122}$$

and multiplying (4.117) by $\dot{\Lambda}(t)$ gives,

$$\dot{\Lambda}(t)\hat{y}(t) \qquad \dot{\Lambda}(t)\Lambda(t)\hat{y}(t) \tag{4.123}$$

By subtracting (4.123) from (4.122) and solving for $\dot{\hat{y}}(t)$ yields,

$$\dot{\hat{y}}(t) = \Lambda^{-1}(t)\dot{\Lambda}(t)\hat{y}(t) + \Lambda(t)\phi(t)x(t). \tag{4.124}$$

The last equation is the matrix counterpart of (4.102). Input output relationship of the estimator in Figure 4.13 can be written as

$$\dot{\hat{y}}(t) = \Phi(t)\hat{y}(t) + \Gamma(t)x(t) \tag{4.125}$$

$$\hat{z}(t) = C^T\hat{y}(t) \tag{4.126}$$

where

$$\Phi(t) = \Lambda^{-1}(t)\dot{\Lambda}(t) \tag{4.127}$$

$$\Gamma(t) = \Lambda(t)\phi(t). \tag{4.128}$$

Equation (4.125) and (4.126) form the state space equivalence of the NC-SONF that can generate the estimates of the signal $s(t)$ from the observed noisy signal, $x(t) = s(t) + n(t)$

Assuming that $\Phi(t)$ can be put in the form,

$$\Phi(t) = A(t) - K(t)C^T \tag{4.129}$$

and letting

$$I(t) = K(t), \tag{4.130}$$

then (4.125) can be written as

$$\dot{\hat{y}}(t) = A(t)\hat{y}(t) + K(t)\left[x(t) - C^T y(t)\right]. \tag{4.131}$$

Referring to the notational differences in Table 4.2, it is evident that (4.131) looks like a Kalman Filter described by (4.93), where $K(t)$ represents the Kalman Gain and $A(t)$ represents the system or signal dynamics, i.e., $\dot{y}(t) = A(t)y(t)$.

Defining the error in the estimation of the states $y(t)$ as

$$\tilde{y}(t) = y(t) - \hat{y}(t) \tag{4.132}$$

allows us to write the differential equation governing the error

$$\dot{\tilde{y}}(t) = \left(A(t) - K(t)C^T\right)\tilde{y}(t) - K(t)n(t) \tag{4.133}$$

where $n(t)$ is the noise in the observed signal $x(t)$. Let $\eta(t) = E[\tilde{y}(t)\tilde{y}^T(t)]$, then the matrix differential equation governing the covariance of the estimation error can now be written as [71]

$$\dot{\eta}(t) = A(t)\eta(t) + \eta(t)A^T(t) - \eta(t)C^T N_o^{-1} C\eta(t) \tag{4.134}$$

and

$$K(t) = \eta(t)C^T N_o^{-1} \tag{4.135}$$

In the above discussion, we have assumed that $\Phi(t)$ can be put in the form of (4.129) and that the Kalman gain is described by (4.130). Thus, the above analysis is only valid if this can be proven. Unfortunately, the mathematical complexities involved in the matrix manipulation form a great obstacle to prove this analytically, as was done for the first order case. We, therefore, resort to simulations to verify (4.131), (4.134) and (4.135) using

$\Lambda(t)$ generated from (4.129). If this can be shown to be true, then it is easy to see that the Kalman gain defined by (4.135) is equivalent to (4.130) and the Kalman equivalence of the NC-SONF is valid.

Kalman Filter		SONF	
$x(t)$	State vector	$x_1(t)$	State vector
$y(t)$	Output signal	$s(t)$	Output signal
$z(t)$	Observed signal	$z(t)$	Observed signal
$F(t)$		$\Lambda(t)$	
$G(t)$		0	
$H(t)$		1	
$P(t)$		$\eta(t)$	
$R(t)$		$N_0/4$	
$K(t)$	$= P(t)H^T(t)R^{-1}(t)$	$K(t)$	$= \Lambda(t)\eta(t)$

Table 4.2 Notational differences in the Kalman and NC-SONF

We use simulation results to verify the Kalman equivalence of a second order SONF. That is, we implement (4.131), (4.134) and (4.135) using $\Lambda(t)$ generated from (4.126) - (4.128) and (4.129) i.e., $\Lambda(t) = \Phi(t) + K(t)C^T - \Lambda^{-1}(t)\Lambda(t) + \Lambda(t)\phi(t)C^T$. The message signal is chosen to be $z(t) = s(t) + A_0\cos(2\pi f_0 t + \theta_0)$, where the normalized frequency $f_0 = 0.2\ Hz$ and θ_0 is a random initial phase that is uniformly distributed on $[0, 2\pi]$. The amplitude A_0 and noise power N_0 are chosen to satisfy $SNR = 3$ dB. Simulation is performed in continuous-time using Gear integration technique [74]. Figure 4.11 shows the actual and the estimated signal $z(t)$. Clearly, using the assumed Kalman equivalence of the NC-SONF, we are able to estimate $s(t)$ adequately. This indicates that the assumption of (4.129) to be correct. Although this result is only an example, in the absence of analytical results, it does provide some validity to the claim of the Kalman equivalence of our approach.

By virtue of the above analytical and numerical results, we can somewhat justify the claim that the SONF filter is equivalent to the Kalman Filter, or in other words that it is in fact, another implementation of the Kalman Solution. With regard to this statement, one important point needs to be addressed. It is well known that the Kalman Filter is an unbiased estimator [75]. The SONF approach with optimal post IMF scaling functions,

on the other hand, is only asymptotically unbiased as can be seen by substituting (4.21) into (4.22) for the coherent case,

$$E[e_{\lambda_{opt}}(t)] = E[V]\left[\frac{c(t)}{c(t)+1/SNR}\right]\phi(t) \neq 0. \tag{4.136}$$

and

$$\lim_{t\to\infty} E[e_{\lambda_{opt}}(t)] = 0. \tag{4.137}$$

Examining (4.136), we can see that since $c(t)$ rapidly increases to a large value (it is the energy of the composing function for time interval t) relative to 1/SNR, the SONF can be said to be an approximately unbiased estimator. Furthermore, for the weak noise case, SONF can be considered an unbiased estimator. If, therefore, the first few output samples of the estimator are discarded, then, for all practical purpose, the SONF can be considered to be an unbiased estimator. Using the analysis in this chapter, in conjunction with the above constraint, we can consider the SONF approach to be a new form of implementing the Kalman Filter.

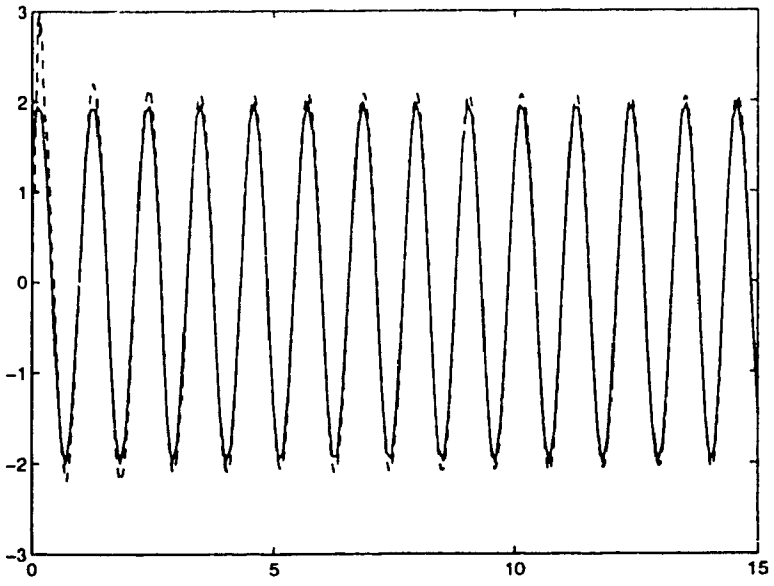


Figure 4.14: Signal estimation via the Kalman equivalent form of the NC-SONF. — Actual Signal $s(t)$; -- Estimated signal $\hat{s}(t)$.

Observe that the variance equation (4.96) in the Kalman Filter does not depend on the input or received data. Therefore, $P(n)$ may be precomputed before any input data is received and used to compute the gains in the optimum filter. This, of course, is equally

true for the SONF which depends on the post-IMF scaling functions that are computed independently of the received data.

In order to solve for the gains in the Kalman Filter, it is necessary to solve a nonlinear matrix differential equation of the Ricatti type (see (4.96)). With the SONF procedure, on the other hand, we do not have such an equation to deal with. The SONF method provides a simple algebraic equation in lieu of the nonlinear differential equation of the Kalman Filter. Consequently, our new approach offers a significantly simpler implementation of the Kalman Filter.

4.8 Conclusion

In this chapter, we have considered processing transient or short-duration signals using a nonparametric approach. The main difficulties in using conventional parametric filtering arise due to the short-duration nature of the such signals. In the literature considerable effort has been devoted to the detection of transients. We have, on the other hand, considered the problem of filtering these signals in the context of enhancement/suppression.

It is well known that MFs increase the output SNR (SNR_o) in a very short time, however, they are exclusively used for the detection purpose. Knowing this, we have combined a special Instantaneous MF (IMF) (maximum SNR_o at each instant of time) with the LS optimization criteria to develop a statistically optimal approach to process short-duration signals. This approach has been termed Statistically-Optimal Null Filter (SONF).

At first the SONF has been developed for the simple case of one term series expansion (coherent case) of the signal under consideration and then later generalized to include an N term expansion (noncoherent case). Such a filter is only locally optimal (each term is optimized individually with no consideration to the other $N-1$ terms). Furthermore, this approach requires that the basis functions in the series expansion be orthogonal for any time interval. This is clearly impossible. Thus, to address this problem, we have derived a globally optimal SONF in which the assumption of orthogonality of the basis functions is removed; as there exists an implicit orthogonalization. The correlation between terms is internally minimized. To implement this globally optimal SONF, SNR of each individual

signal component is required. This information is generally not *a priori* known, as such this approach is only of theoretical value. We address this issue by developing a recursive SONF in which minimal *a priori* knowledge is required. Often, when filtering exponentially damped sinusoids, the knowledge of the damping factor is not *a priori* known. This has also been considered in this chapter by using the SONF in an adaptive configuration.

We have derived a time-varying nonparametric method of processing signals that is optimal under the MMSE criterion. With such an optimality criterion, it was anticipated that our approach may be equivalent to the well known Kalman Filter. We have shown this to be in fact true under a weak constraint. The Kalman Filter is an unbiased estimator while the SONF provides only asymptotically unbiased filtering. This, however, has been shown to be a minor hindrance, as the SONF for all practical purpose can be considered to be an unbiased estimator. If the first few output samples of the SONF are discarded, then the SONF can be said to equivalent to the Kalman Filter and all the properties of the Kalman Filter are applicable to the SONF.

4.9 Appendix

The gradient of $e(n)$ w.r.t. β is (see (4.82))

$$\Psi(n) =) \frac{\partial \lambda^T(n)}{\partial \beta} \nu(n) + \lambda^T(n) \frac{\partial \nu(n)}{\partial \beta}. \tag{4.138}$$

The derivatives on the right hand side of (4.138) are defined as

$$\frac{\partial \lambda^T(n)}{\partial \beta} = \begin{bmatrix} \frac{\partial \lambda_1(n)}{\partial \beta} \\ \frac{\partial \lambda_2(n)}{\partial \beta} \end{bmatrix}^T \quad \text{and} \quad \frac{\partial \nu(n)}{\partial \beta} = \begin{bmatrix} \frac{\partial \nu_1(n)}{\partial \beta} \\ \frac{\partial \nu_2(n)}{\partial \beta} \end{bmatrix}^T. \tag{4.139}$$

We now evaluate each of these two derivatives and substitute back in (4.138) to find $\Psi(n)$. Since each component in $\lambda(n)$ is defined as

$$\lambda_i(n) = \frac{\phi_i(n)}{\| \phi_i(n) \|^2}, \tag{4.140}$$

we first evaluate the derivatives of $\lambda_i(n)$ w.r.t. β,

$$\frac{\partial \lambda_i(n)}{\partial \beta} = \frac{\frac{\partial \phi_i(n)}{\partial \beta} \cdot \| \phi_i(n) \|^2 - \phi_i(n) \frac{\partial \| \phi_i(n) \|^2}{\partial \beta}}{\| \phi_i(n) \|^4} \tag{4.141}$$

Let $\phi_i(n) = e^{\beta n} \phi_i'(n)$. The derivative of $\phi_i(n)$ w.r.t. β is

$$\frac{\partial \phi_i(n)}{\partial \beta} = n e^{\beta n} \phi_i'(n) = n \phi_i(n). \tag{4.142}$$

Similarly,

$$\frac{\partial \| \phi_i(n) \|^2}{\partial \beta} = \sum_m^n 2m \phi_i^2(m). \tag{4.143}$$

Substituting (4.142) and (4.143) back in (4.141) yields,

$$\frac{\partial \lambda_i(n)}{\partial \beta} = \frac{n \phi_i(n)}{\partial \| \phi_i(n) \|^2} - \frac{2 \phi_i(n) \sum_m^n m \phi_i^2(m)}{\sum_m^n \phi_i^2(m)} \tag{4.144}$$

The derivative of $\nu_i(n)$ w.r.t. β is

$$\frac{\partial \nu_i(n)}{\partial \beta} = \sum_m^n x(n) \frac{\partial \phi_i(m)}{\partial \beta} \tag{4.145}$$

$$= \sum_m^n m x(m) \phi_i(m). \tag{4.146}$$

Letting,

$$A_i(n) = \sum_m^n \phi_i^2(m) = \sum_m^{n-1} \phi_i^2(m) + \phi_i^2(n)$$

$$= A_i(n-1) + \phi_i^2(n) \tag{4.147}$$

$$B_i(n) = \sum_m^n m\phi_i^2(m) = \sum_m^{n-1} m\phi_i^2(m) + n\phi_i^2(n)$$

$$= B_i(n-1) + n\phi_i^2(n) \tag{4.148}$$

$$C_i(n) = \sum_m^n mx(m)\phi_i(m) = \sum_m^{n-1} mx(m)\phi_i(m) + nx(n)\phi_i(n)$$

$$= C_i(n-1) + nx(n)\phi_i(n) \tag{4.149}$$

we can write each of the vector components of (4.139) as

$$\frac{\partial \lambda_i}{\partial \beta} = \frac{n\phi_i(n)}{A_i(n)} - 2\phi_i(n)\frac{B_i(n)}{A_i^2(n)} \tag{4.150}$$

$$\frac{\partial \nu_i(n)}{\partial \beta} = C_i(n). \tag{4.151}$$

Substituting (4.150), (4.151) in (4.138), $\Psi(n)$ can be written as,

$$\Psi(n) = \begin{bmatrix} \frac{\phi_1(n)}{A_1(n)}\left(n - 2\frac{B_1(n)}{A_1(n)}\right) \\ \frac{\phi_2(n)}{A_2(n)}\left(n - 2\frac{B_2(n)}{A_2(n)}\right) \end{bmatrix}^T \begin{bmatrix} \nu_1(n) \\ \nu_2(n) \end{bmatrix} + \begin{bmatrix} \frac{\phi_1(n)}{A_1(n)} \\ \frac{\phi_2(n)}{A_2(n)} \end{bmatrix}^T \begin{bmatrix} C_1(n) \\ C_2(n) \end{bmatrix} \tag{4.152}$$

$$= \sum_{i=1}^{2} \frac{\phi_i(n)}{A_i(n)}\left[n\nu_i(n) - 2\frac{\nu_i(n)}{A_i(n)}B_i(n) + C_i(n)\right] \tag{4.153}$$

Chapter 5

Statistically Optimal Null Filters: Simulation Results and Applications

In the previous chapter, we described a new non-parametric approach for suppression/enhancement of signals of known/unknown waveform shapes. Several variants of the basic SONF approach were presented. In this section, we compare the effectiveness and performance of these new methods via simulations and show that they can indeed be used for the suppression/enhancement of short record signals. We compare their error performance through Monte Carlo simulations. All simulations are done in discrete-time.

In Section 5.1, we present the results of processing stationary signals by the SONFs, while in following section we give two examples of processing a time-varying signal. The latter is meant to only indicate the feasibility of using the SONFs for these type of signals and is not an exhaustive evaluation. In these two sections, we also compare our new approach to the conventional parametric CNFs of [26]. Section 5.3 considers the application of SONFs to separate highly correlated signals with overlapping Fourier Spectra. For processing damped sinusoids with unknown damping coefficients, simulations results of the adaptive SONF are presented in Section 5.4.

5.1 Stationary Signal

Example 5.1 This example is used to show the viability of the basic coherent (Figure 4.6) and the noncoherent (Figure 4.7) SONFs. We test its functionality, that it can indeed be used for the narrowband (NB) signal suppression/enhance-ment problem. Since the SNR is generally assumed to be unknown, we use the suboptimal post-IMF scaling function, $\lambda(n)$.

Let the input consist of an additive mixture of a sinusoid, $s(n) = cos(\omega_o n + \theta_o)$ and a zero-mean white gaussian noise, $n(n)$. The input SNR is -3 dB. The sinusoidal frequency ω_o and the initial phase θ_o are assumed to be known. For the noncoherent case, the initial phase information is not used. Here, we point out that the NC-SONF uses the same amount of *a priori* information as the parametric methods of Chapter 3.

The composing or the basis functions for the coherent implementation of the SONF can be written as $\phi^c(n) = cos(\omega_o n + \theta_o)$, where the superscript 'c' indicates coherent. The

sig...l $s(n)$ can also be written as

$$s(n) = A\cos(\omega_o n + \theta_o) = v_1\cos(\omega_o n) + v_2 sin(\omega_o n)$$
$$= v_1\phi_1^{nc}(n) + v_2\phi_2^{nc}(n) \quad (5.1)$$

where $\phi_1^{nc}(n)$ and $\phi_2^{nc}(n)$ are orthogonal functions (note that this orthogonality is true only over the intervals kT_o where k is any integer and T_o is the period of $s(n)$) that serve as the basis functions for the NC-SONF. The coefficients v_1 and v_2 are unknown since A and θ_o are unknown. Our earlier assumption of orthogonality over all t is not true, we nevertheless, observe promising results. Figure 5.1 show the results of suppressing the sinusoid by the two approaches. It is clear that both forms of the SONF provide effective suppression of the sinusoid. Figure 5.2 show the MSE in suppressing $s(n)$. MSE

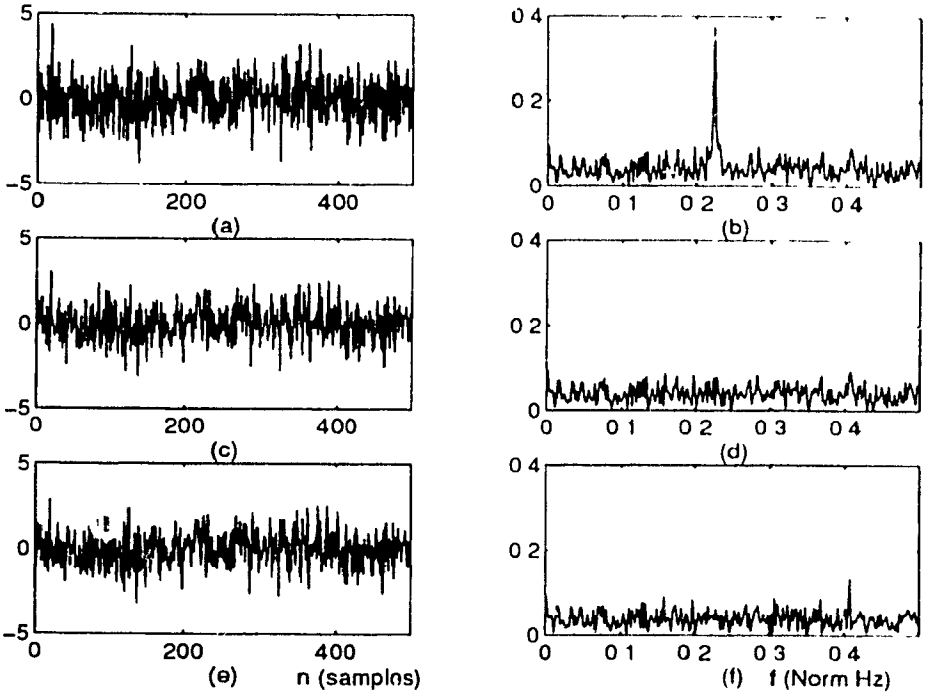


Figure 5.1: Results of Filtering via SONF and NC-SONF.(a) Input mixture $x(n)$; (b) PSD of $x(n)$; (c) Output of SONF, $\hat{n}(n)$; (c) PSD of output in (c); (e) Output of NC SONF, $\hat{n}(n)$; (f) PSD of output in (e).

was calculated based on an ensemble average of 500 trials. The initial phase of the input sinusoid was chosen to be uniformly distributed over [0,2π]. The NC-SONF gives errors that are 3 dB poorer than the coherent SONF. This, of course, is in good correspondence

to coherent and noncoherent detection based on classical MFs used in communications.

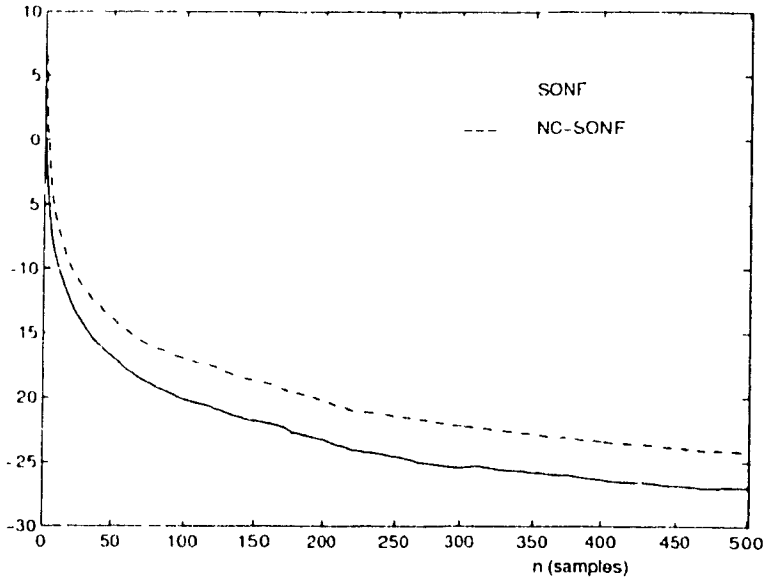


Figure 5.2. MSE in suppressing a sinusoid by the SONF and the NC SONF.

Example 5.2 In this example we compare the performance of the optimal and the sub-op' post-IMF scaling function in the SONF of Figure 4.6. The objective is to show the effect of using additional information of the SNR. The input $x(n)$ is the same as in the last example. Figure 5.3 shows the MSE in suppressing $s(n)$ for the first 150 samples. It is clear that the knowledge of the additional information of the SNR gives a lower MSE at the start of the filtering as predicted by (4.26). This improvement is only valid at the very start of the algorithm - about 50-100 samples. As time progresses, the two approaches yield equivalent performance. Asymptotically, the two methods are equivalent.

It should be noted that the optimal approach allows for a stable implementation of the filter as there can be no division by zero in the post-IMF scaling function. The use of SNR avoids a possible division by zero. However, it should also be noted that the additional information of the SNR is generally not available. This is particularly so in the noncoherent implementations where individual component SNRs would be required.

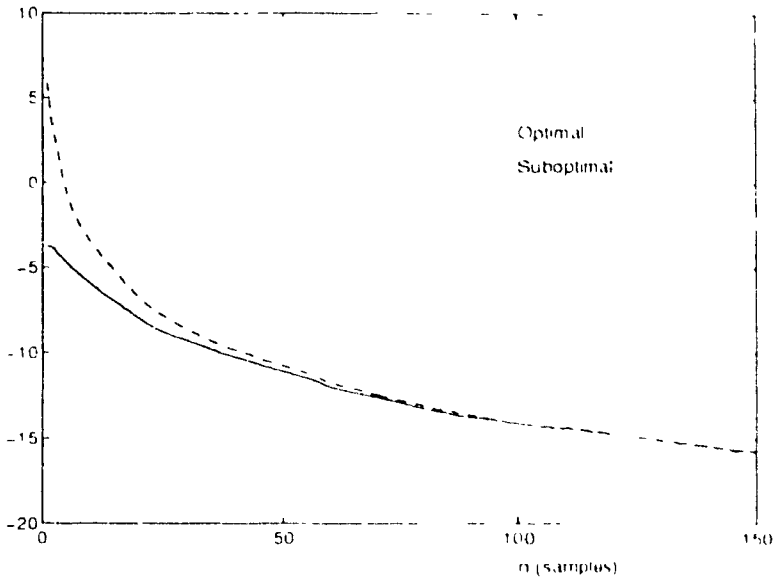


Figure 5.3: MSE in suppressing $s(n)$ by using the optimal and suboptimal post IMF scaling function in the coherent SONF.

Example 5.3 In this example, we study the effects of using the sliding GS orthogonalization in SONF of Figure 4.7 for the case of $N = 2$. We use the suboptimal post IMF scaling function and all results are based on 500 run Monte Carlo simulations. The input consist of an additive mixture of noise and the signal $s(n)$,

$$s(n) = A_1 cos(\omega_1 n + \theta_1) + A_2 cos(\omega_2 n + \theta_2)$$

where the normalized frequencies are $\omega_1 = 0.5$ and $\omega_2 = 0.6$ Hz and the initial phases, θ_1 and θ_2 are uniformly distributed over $[0, 2\pi]$. Figure 5.4 shows the MSE for estimating $s(n)$ with/without using sliding GS-orthogonalization in the NC SONF. At the start of filtering, the performance of the NC-SONF with sliding GS-orthogonalization is better than without it. As time progresses, the performance of the SONF without the orthogonalization approaches that of the one with it. These results indicate that given signals of sufficiently long duration, we do not need to use the orthogonalization procedure. However, for short-duration signals it is desirable.

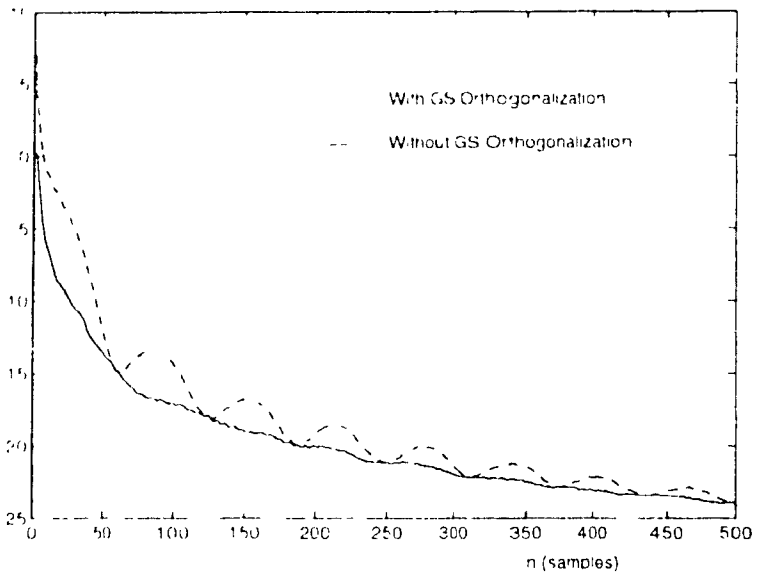


Figure 5.4: Comparisons of the MSE in suppressing $s(n)$ with/without the GS orthogonalization

Example 5.4 In this example, we compare the differences in the recursive SONF and NC SONF with sliding GS-orthogonalization. The input is the same as in Example 5.3 except the initial phase θ_o is not known. Figure 5.5 shows the results of the filtering. Both approaches are able to adequately suppress $s(n)$. Figure 5.6 shows the MSE in suppressing/enhancing $s(n)$. The results are from an ensemble average of 500 trials. Both methods show equivalent performance. Given that the computation requirements for the recursive method are significantly lower than the sliding GS orthogonalization, it is the preferred one for implementation.

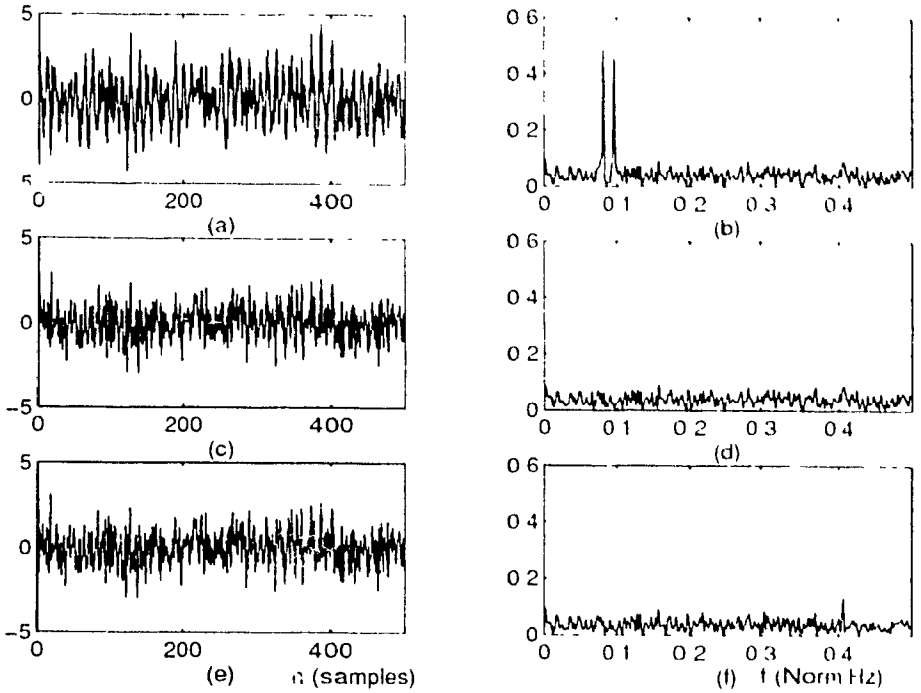


Figure 5.5: Results of Filtering via the recursive SONF and GS orthogonalized NC SONF.(a) Input mixture $x(n)$; (b) PSD of $x(n)$; (c) Output of recursive SONF, $u(n)$. (c) PSD of output in (c); (e) Output of GS orthogonalized NC SONF, $u(n)$; (f) PSD of output in (e).

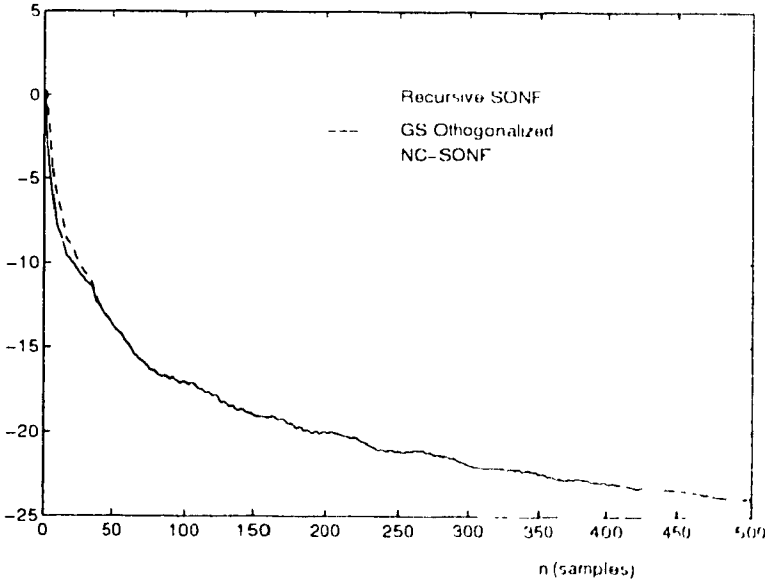


Figure 5.6: MSE in suppressing $s(n)$ by using the recursive SONF and the GS orthogonalized NC-SONF

Example 5.5 In this example we compare the recursive SONF against the parametric methods of the last chapter. The MSE curves of the two approaches are shown in Figure 5.7. Clearly, after the initial transient period, the recursive SONF proves to be significantly better than the CNF. It is also significantly better than the 4th order LMS Optimal Modified Notch Filter (MNF) of Chapter three. For both of the parametric filters we have used the debiasing coefficient $\alpha = 0.9$. Moreover, the recursive SONF is consistent

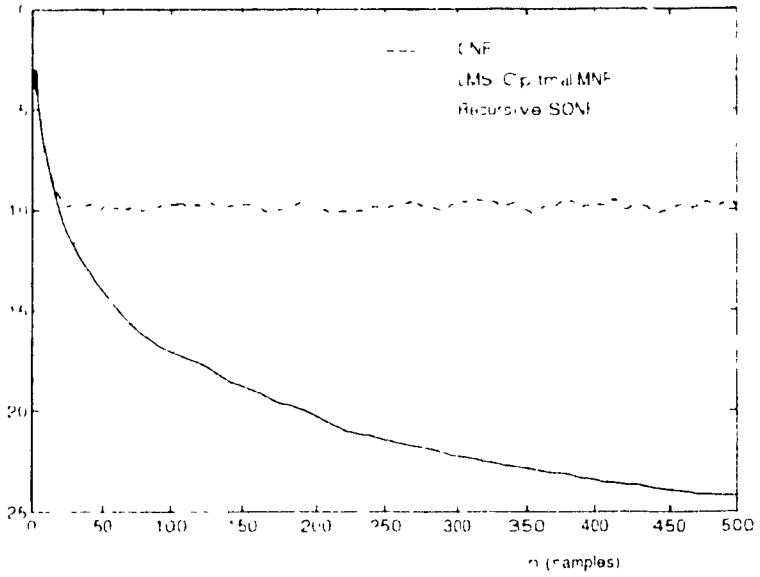


Figure 5.7: MSE in suppressing $s(n)$ by using the recursive SONF, CNF and the LMS Optimal MNF.

Example 5.6 All examples thus far considered demonstrate the ability of the newly developed SONF to suppress enhance NB signals immersed in WGN. From the various MSE curves it is clear that the error drops to a low value very quickly indicating its ability to process signals with a short transient duration. Example 5.5 shows that the SONF is significantly better than the conventional parametric CNF with a debiasing coefficient $\alpha = 0.9$ (recall that a low debiasing coefficient is required to reduce the transient duration).

In this example we show that the SONF can indeed be used to process transient or short duration signals. We also compare the performance of SONF against the conventional CNF.

The input consists of a mixture of a damped sinusoid in zero mean WGN

$$x(n) = s(n) + w(n), \qquad n = 1, 2, \ldots, N \tag{5.4}$$

where $s(n) = Ae^{-\alpha n}\cos(\omega_0 n + \theta)$. The damping coefficient α and the sinusoidal frequency ω_0 are *a priori* known, while the initial phase θ and the amplitude A are unknown. The SNR is approximately -10 dB. For the damped sinusoid in noise case we define the SNR as,

$$SNR = \frac{1}{N}\frac{\sum_{n=1}^{M} s^2(n)}{V} \tag{5.5}$$

where $M \leq N$ is the time required for the amplitude of the damped sinusoid to decay to 10% of its maximum value and V is the power of the noise. We use the recursive SONF and the conventional CNF. The knowledge of the damping factor is incorporated in the calculation of the CNF coefficient. Figure 5.8 shows the results of estimating $s(n)$ via the recursive SONF. The MSE in the estimation via the recursive SONF and the CNF are shown in Figure 5.9. The MSE in the proposed approach drops to a small value very quickly and continues to improve, while for the CNF approach it reaches a saturation and no further improvements are obtainable. The level of the steady state error in the CNF is a function of the debiasing coefficient as was shown in chapter 3. As α is increased the MSE reduces, however, at a cost of longer transient duration. This indicates the superiority of the SONF to process short-duration signals.

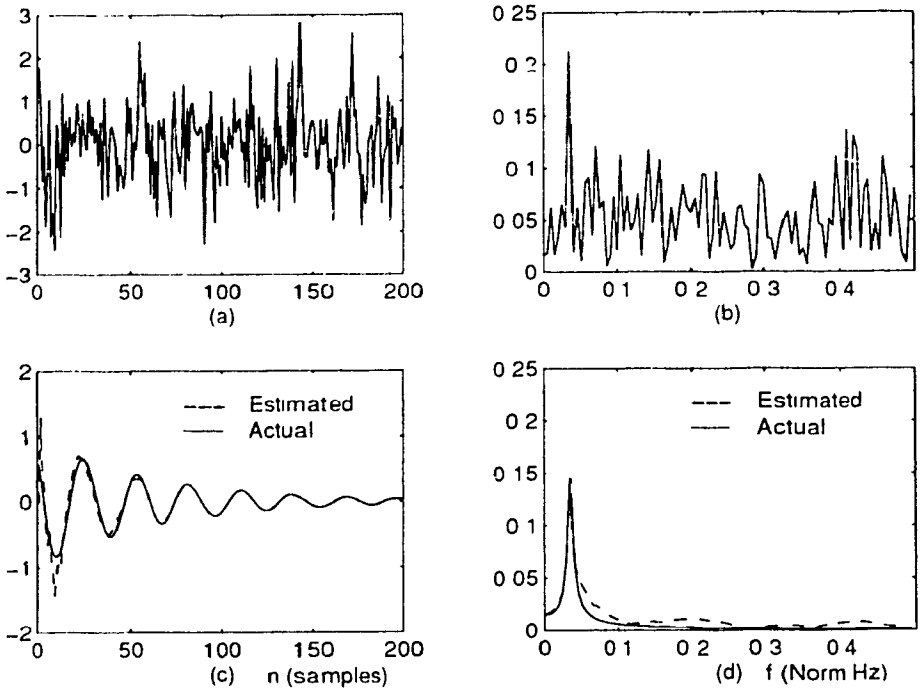


Figure 5.8: Results of estimating a damped sinusoid with the recursive SONF. (a) input mixture $x(n)$; (b) PSD of $x(n)$; (c) True and Estimated signal; (d) PSD of the signals in (c).

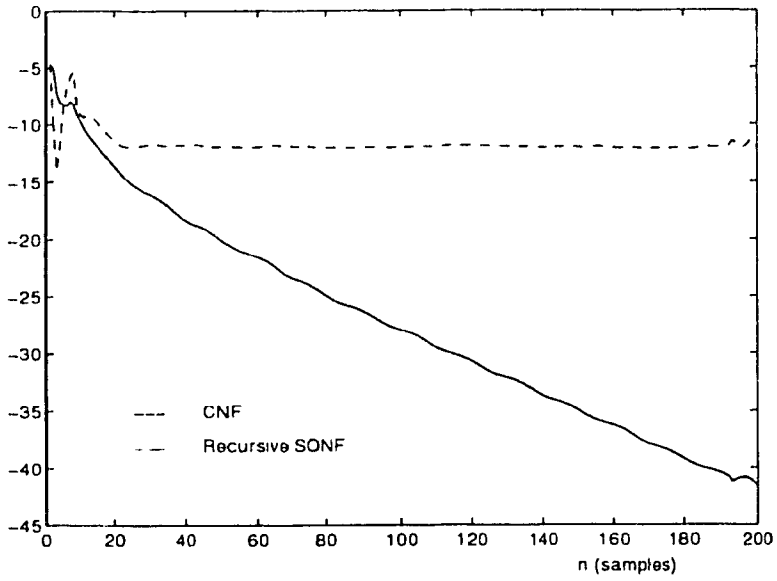


Figure 5.9: MSE in estimating a damped sinusoid by the conventional CNF and the recursive SONF.

5.2 Time-Varying Signals

Generally for the detection of signals, MFs are useful only in the stationary environment. However, since our approach is used for the estimation (not detection) of signals instant by instant, it should be possible to extend this approach for the processing of non stationary signals. This, of course, must be true as this approach has been shown to be equivalent to the Kalman Filter. We present the following two simulation results to validate the use of SONF for the processing of time-varying signals. These example serve to **only indicate the possibilities** of using the recursive SONF for the processing of (FM) non stationary signals. As earlier, all simulations are in discrete-time.

Example 5.7 The signal $s(n)$ in (4.2) is now chosen to be a frequency modulated type. It can be written as $s(n) = Acos(\varphi(n) + \theta_o)$ where $\varphi(n) = \omega_c n + \sum_m^n \omega(m)$. The modulating frequency $\omega(n)$ is generated by passing zero-mean white noise sequence through an FIR lowpass filter with a cutoff frequency $\omega_{cut} = .1 rad/s$. $n(n)$ is zero-mean white noise uncorrelated with the one used to generate $\omega(n)$. The SNR is -3 dB and the instantaneous phase $\varphi(n)$ is *a priori* known. We estimate the FM signal using the recursive SONF where the basis functions are set to be $cos\varphi(n)$ and $sin\varphi(n)$. Figure 5.10 shows the results of estimating $s(n)$ via the recursive SONF and the Almost-Symmetrical Time Varying ARMA (ASTV-ARMA) of [32] for the debiasing coefficient $\alpha = 0.9$. Figure 5.11 shows the MSE in the estimate of $s(n)$ by the two methods. For comparison purpose, we have included ASTV-ARMA results for $\alpha = 0.99$ and $\alpha = 0.995$. The performance of the ASTV-ARMA model starts to approach that of the SONF as α is increased, however, the transient duration starts to become excessively long. This curtails the usage of the ASTV-ARMA for transient or short-duration signals. Clearly, the SONF approach provides superior results for short-duration signals.

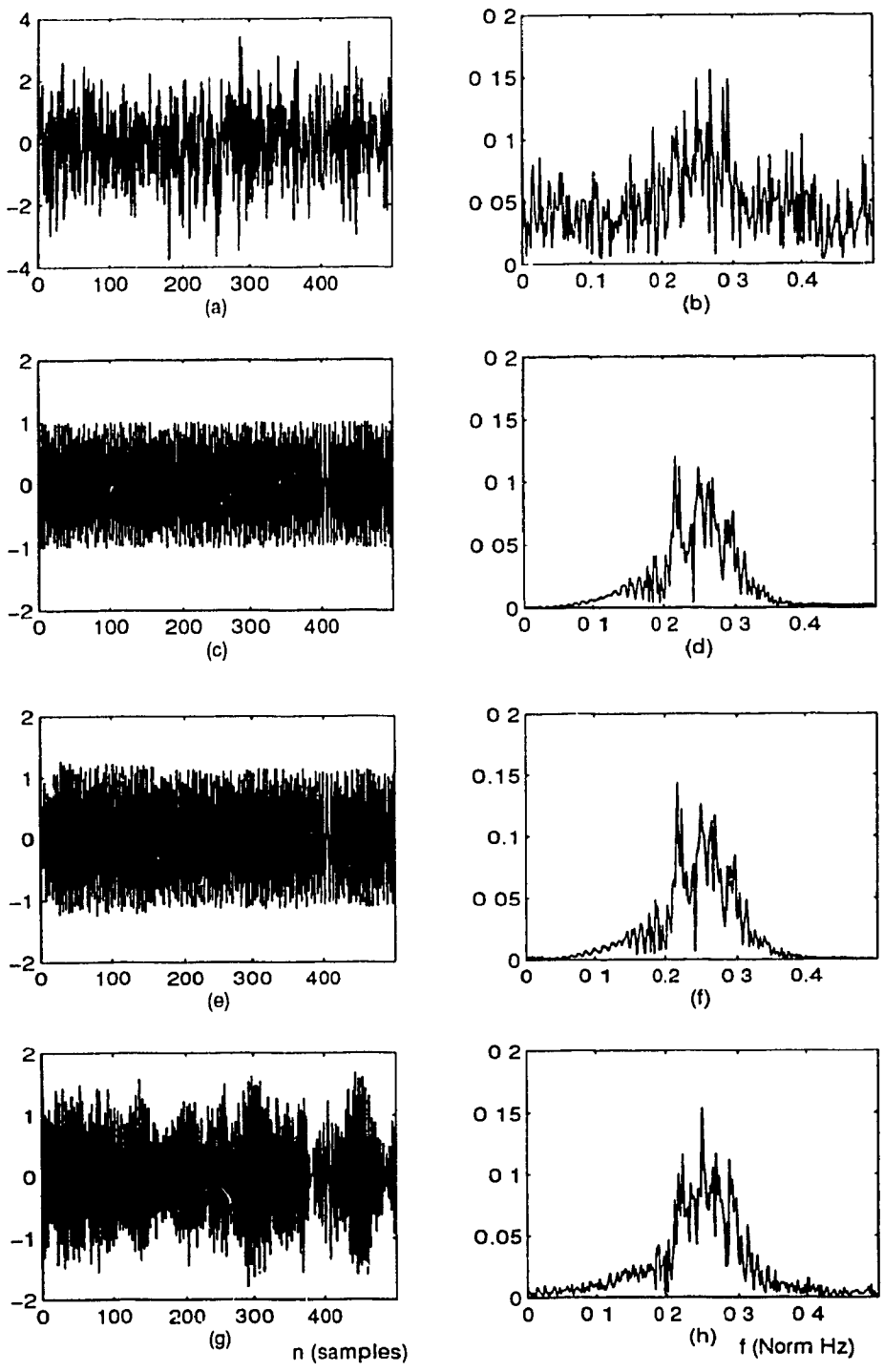


Figure 5.10: Results of extracting a FM signal in the presence of noise via the SONF and the ASTV-ARMA with $\alpha = .9$. (a) Input mixture $x(n)$; (b) PSD of $x(n)$; (c) $s(n)$; (d) PSD of $s(n)$ in (c); (e) Output of SONF, $\hat{s}(n)$; (f) PSD of $\hat{s}(n)$ in (e); (g) Output of ASTV-ARMA, $\hat{s}(n)$; (h) PSD of $\hat{s}(n)$ in (g).

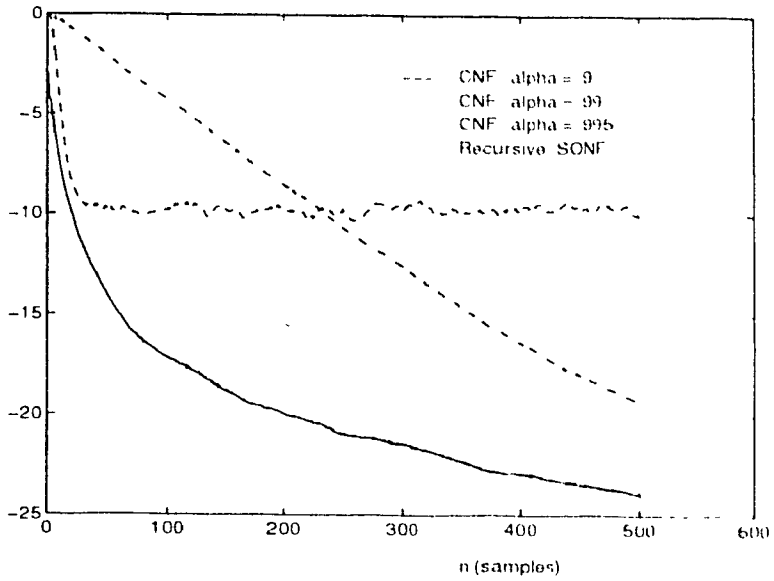


Figure 5.11: MSE in estimating an FM signal in the presence of noise

Example 5.8 In this example, we show the separation of two finite duration FM signals with overlapping Fourier Spectra. The two FM signals are generated in a manner similar to that in Example 5.7. Figure 5.12 shows the results of separating via recursive SONF and the ASTV-ARMA with a debiasing coefficient $\alpha = 0.9$. Figure 5.13 shows the MSE in estimating $s_1(t)$, one of the two FM signals in the input mixture, for the two methods. Clearly, our proposed approach provides improved results.

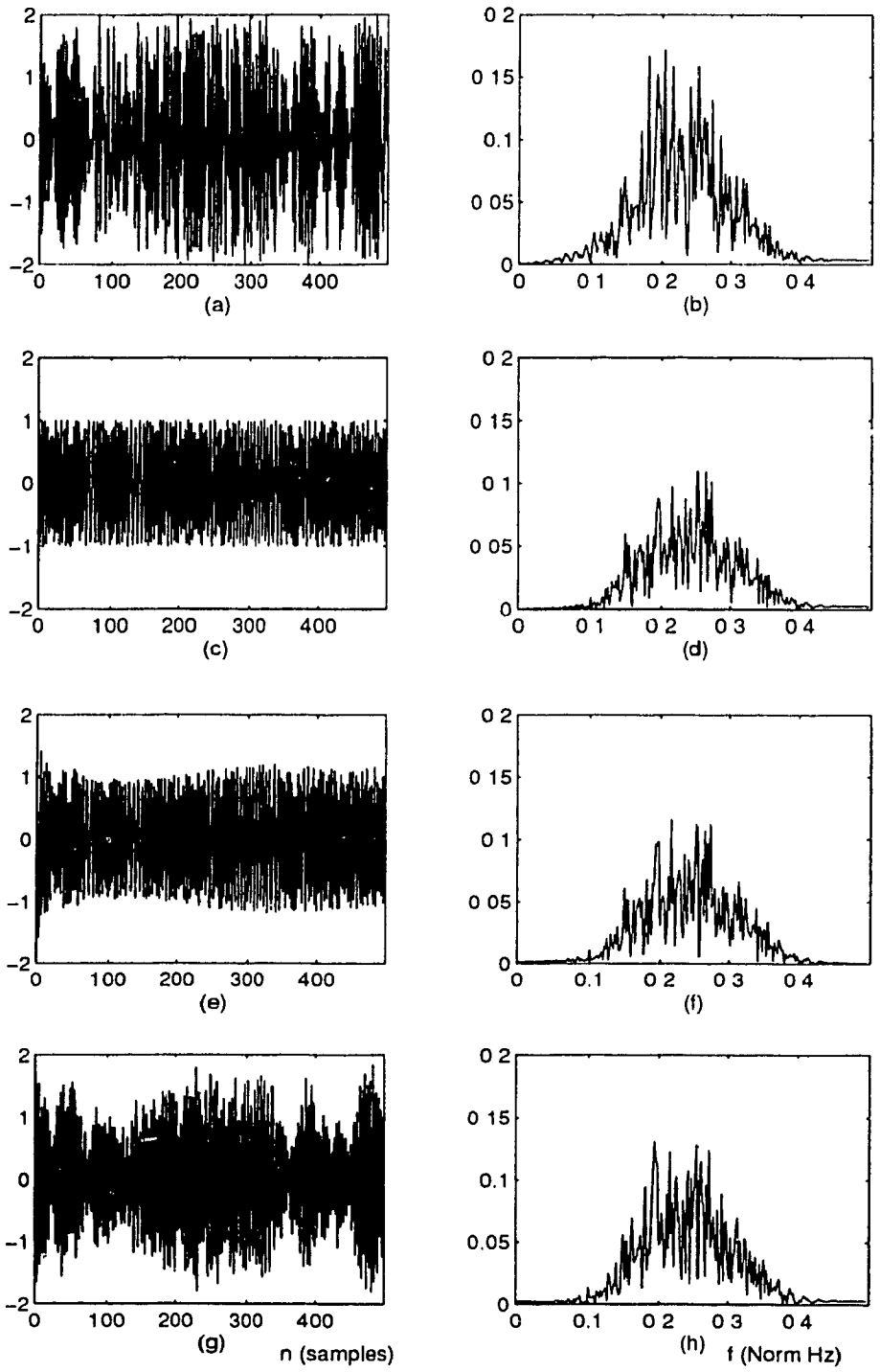


Figure 5.12: Results of separating two FM signals via SONF.(a) Input mixture $x(n)$; (b) PSD of $x(n)$; (c) $s_1(n)$, the first FM signal; (d) PSD of $s_1(n)$ in (c); (e)Estimate via SONF, $\hat{s}_1(n)$; (f) PSD of $\hat{s}_1(n)$ in (e); (g) Estimate via ASTV-ARMA, $\hat{s}_1(n)$; (h) PSD of $\hat{s}_1(n)$ in (g).

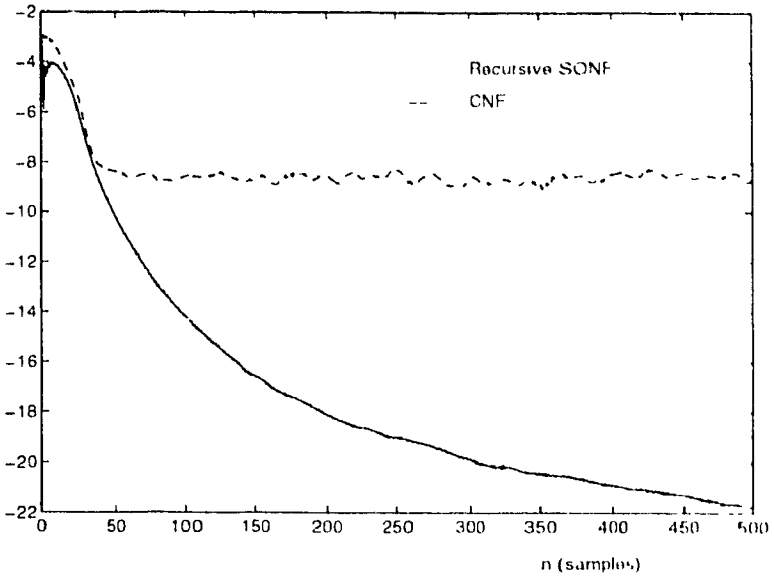


Figure 5.13: MSE in separating two FM signals.

5.3 Application: Separation of Signals in AWGN [1]

Here we consider the application of separating closely spaced NB signals immersed in white gaussian noise using the newly developed SONF. We will show the performance of the SONF for separating closely spaced damped sinusoids (transient signals) in noise. We consider the case of two damped sinusoids in noise.

Let the input $x(t)$ be composed of a mixture of two NB signal components in AWGN,

$$\begin{aligned} x(t) &= s_1(t) + s_2(t) + n(t) \\ &= v_1\phi_1(t) + v_2\phi_2(t) + n(t), \qquad t \in [0, T] \end{aligned} \tag{5.4}$$

where the composing functions $\phi_1(t)$ and $\phi_2(t)$ are *a priori* known and $n(t)$ is zero-mean white gaussian noise. The composing functions are not orthogonal for all $t \in [0, T]$, i.e., (4.27) does not hold. By using the sliding GS orthogonalization procedure, the input mixture can be written in terms of new orthogonal functions, $\psi_1(t)$ and $\psi_2(t)$,

$$\begin{aligned} x(t) &= A_1(t)\psi_1(t) + A_2(t)\psi_2(t) + n(t) \\ &= x_1(t) + x_2(t) + n(t), \quad t \in [0, T]. \end{aligned} \tag{5.5}$$

We have converted the case of constant coefficients v_1 and v_2 in (5.4) to a general case where the coefficients are time varying. Since the SONF approach is LTV, we apply it to separate $x_1(t)$ and $x_2(t)$ from which $s_1(t)$ and $s_2(t)$ can be determined.

Given that the basis functions $\phi_1(t)$ and $\phi_2(t)$ are *a priori* known, then the new orthogonal basis functions $\psi_1(t)$ and $\psi_2(t)$ are also known. By slightly modifying the SONF approach, it is possible to separate the two components, $s_1(t)$ and $s_2(t)$ directly. Since we are assuming that there is no *a priori* knowledge about the SNR of each component, we use the suboptimal post-IMF scaling function.

The output of the IMF part is scaled by $1/\|\psi_i(t)\|_t^2$ to estimate the amplitudes $A_i(t)$,

$$\begin{aligned} \hat{A}_1(t) &= (v_1 - \rho_{21}(t)v_2) + n'_{o1}(t) = A_1(t) + n'_{o1}(t) \\ \hat{A}_2(t) &= v_2 + n'_{o2}(t) = A_2(t) + n'_{o2}(t) \end{aligned} \tag{5.6}$$

If we multiply $\hat{A}_2(t)$ by $\phi_2(t)$ we obtain,

$$\hat{s}_2(t) = \hat{A}_2(t)\phi_2(t) = s_2(t) + n_{02}(t). \tag{5.7}$$

If we now add $\rho_{21}(t)\hat{A}_2(t)$ to $\hat{A}_1(t)$ and multiply the result by $\phi_1(t)$ we obtain,

$$\hat{s}_1(t) = \left[\rho_{21}(t)\hat{A}_2(t) + \hat{A}_1(t)\right]\phi_1(t) = s_1(t) + n_{o1}(t). \tag{5.8}$$

Notice that we did not use $\lambda'_i(t) = \psi_i(t)/\|\psi_i(t)\|_t^2$ directly. Figure 5.14 shows the resulting

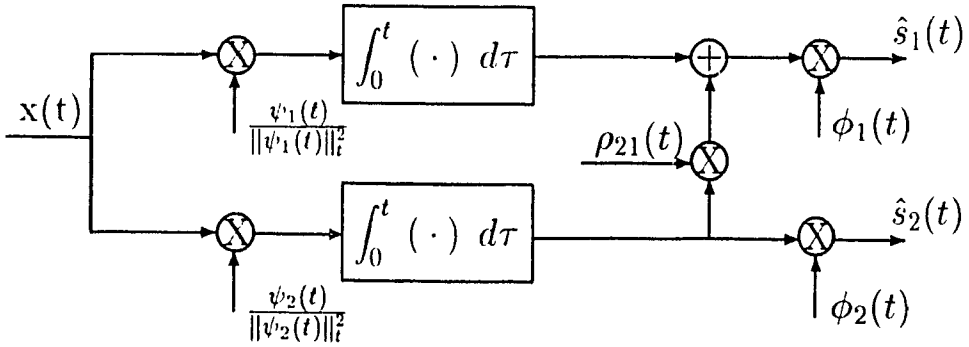


Figure 5.14: Separation of closely spaced signals.

modified SONF circuit to separate the two signals embedded in noise. The output noise

in each of the two components can be written as,

$$n_{o1}(t) = \left[\frac{< n(t), \psi_1(t) >_t}{\|\psi_1(t)\|_t^2} + \rho_{21}(t) \frac{< n(t), \psi_2(t) >_t}{\|\psi_2(t)\|_t^2} \right] \psi_1(t) \tag{5.9}$$

and

$$n_{o2}(t) = \left[\frac{< n(t), \psi_2(t) >_t}{\|\psi_2(t)\|_t^2} \right] \psi_2(t). \tag{5.10}$$

The MSE in estimating each of the two components, $c_1\phi_1(t)$ and $c_2\phi_2(t)$ is given as

$$E[n_{o1}^2(t)] = N_o \phi_1^2 \left[\frac{1}{\|\psi_1(t)\|_t^2} + \rho_{21}^2(t) \frac{1}{\|\psi_2(t)\|_t^2} \right] \tag{5.11}$$

and

$$E[n_{o2}^2(t)] = N_o \frac{\phi_2^2(t)}{\|\psi_2(t)\|_t^2} \tag{5.12}$$

Asymptotically the noise power in each of the two estimated signals approaches zero.

Example 5.9 We now present simulation results [1] describing the effectiveness of using the approach of Figure 5.14 to extract a short duration NB signal component in the presence of heavy noise and an interference strongly correlated with the signal. The input mixture is described as

$$x(n) = s(n) + r(n) + n(n) \tag{5.13}$$

where $s(n) = A_s\phi_s(n)$ is a message signal, $r(n) = A_r\phi_r(n)$ is the interference and $n(n)$ is zero-mean AWGN. The composing functions $\phi_s(n) = e^{\beta_s n}\cos(f_s n)$ and $\phi_r(n) = e^{\beta_r n}\cos(f_r n)$ are *a priori* known with $\beta_s = -0.016$, $\beta_r = -0.012$, and normalized frequencies of $f_s = 0.035$ and $f_r = 0.01$ Hz. $\phi_s(n)$ and $\phi_r(n)$ are highly correlated with a correlation coefficient $\rho_{sr}(n)$ (see Figure 5.15) that approaches unity at time interval $n \to 0$ and decreases to a value of 0.5 towards the end of the considered time interval. The amplitudes A_s, A_r and the noise power are chosen such that the SNR and SIR are each -10 dB. Figure 5.16 describes the results of estimating $s(n)$. This approach provides rather large magnitudes of the first few samples in the estimated signal $s(n)$ due to the division by $\|\psi_2(n)\|_t^2$, which is a rather small number at the onset of filtering. We have, therefore, not used the first four samples in the results of Figure 5.16. It is evident from Figure 5.16c, that after an initial transient period of approximately 25 samples, a very good estimate of $s(n)$ is achieved.

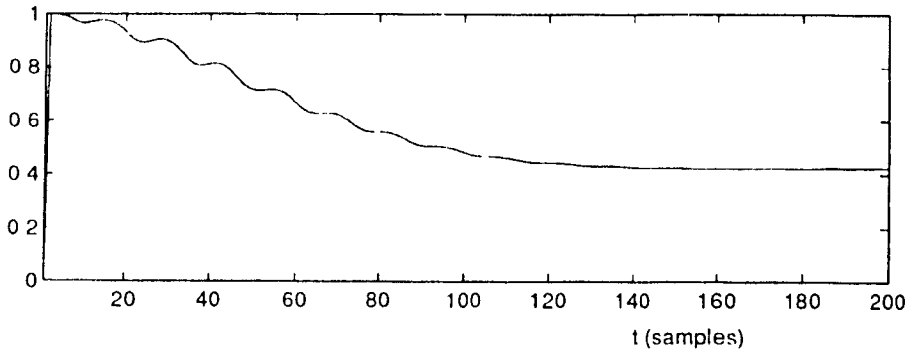


Figure 5.15: Cross correlations coefficient, $\rho_{sI}(n)$, of the interference and the useful signal.

The MSE curves of the modified SONF approach of Figure 5.14 and that of the CNF of [26] are shown in Figure 5.17. The knowledge of the decay factor β_s and frequency f_s has been incorporated in the calculation of the conventional CNF coefficients. Clearly, after the initial transient period, the GS orthogonalized SONF approach provides a much lower MSE.

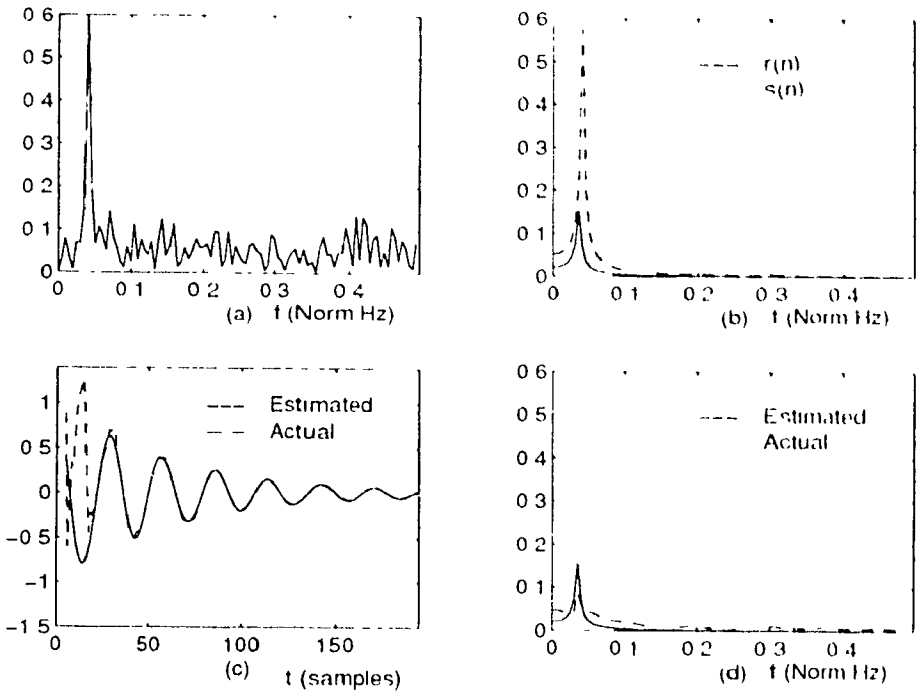


Figure 5.16: Results of estimating a $s(n)$ in the presence of highly correlated interference signal and heavy noise. (a) Input signal $x(n)$, (b) $s(n)$ and $r(n)$; (c) Estimated and true $s(n)$; (d) PSD of estimated $s(n)$

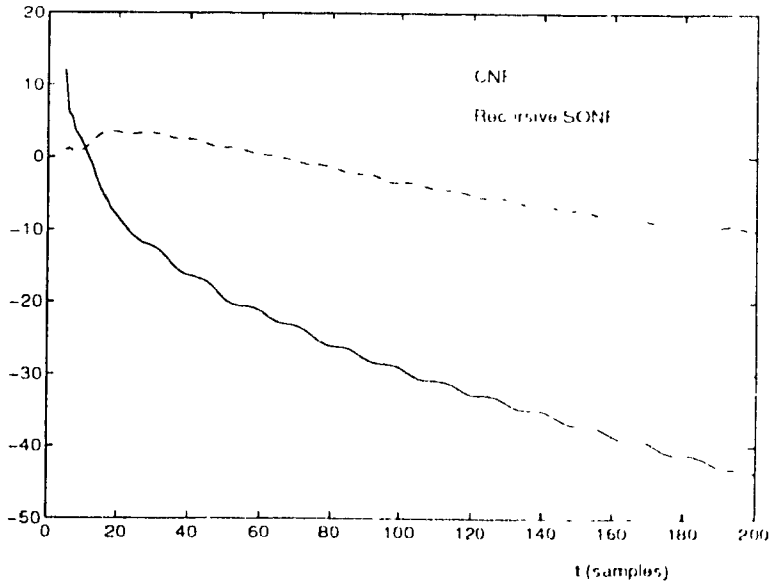


Figure 5.17: MSE in estimating the useful signal $s(n)$ via the SONF and the ASTV ARMA.

5.4 Adaptive SONFs

In this section, we present the simulation results of applying the adaptive SONF for the enhancement/suppression of damped sinusoids with unknown damping coefficients. We present two examples to show the effectiveness of the Adaptive SONFs.

Example 5.10 This example shows the simple case of a single damped sinusoid in zero-mean AWGN. The damped sinusoid $s(n)$ is defined as,

$$s(n) = Ae^{\beta n}cos(\omega_o n + \theta_o) \tag{5.11}$$

where A, θ_o and β are the unknown amplitude, initial phase and damping coefficient, respectively. The sinusoidal frequency ω_o is *a priori* known. In this example we set the damping coefficient to $\beta = -0.0513$ (or $e^{\beta} = 0.95$), initial phase is uniformly distributed on the interval $[0, 2\pi]$ and the amplitude is chosen such that the $SNR = 10dB$. The SNR is defined as in (5.3). Step-size parameter in the adaptation is set to $\gamma = 0.01$.

Figure 5.18 shows the results of extracting $s(n)$ from the noisy data. The hat denotes the estimate of $s(n)$. Note that $s(n)$ represents a very short-duration signal. Its useful (significant) portion is only the first 40-50 samples. In the adaptive estimation of β, during this time interval, a reliable estimate of $\hat{\beta}$ does not exist. As such, we do not expect the output of the filter to represent $s(n)$ with any fidelity (i.e., we are using SONF with incorrect basis functions). To combat this, in the estimate of $s(n)$ in Figure 5.18, we have applied the idea of *a posteriori* filtering [76]. With this method, the idea is to simply refilter the complete data once the filter parameters have been estimated. In our case, once the damping coefficient β has been estimated with sufficient accuracy, we can generate the necessary basis functions ($\phi_1(n)$ & $\phi_2(n)$) to implement SONF. Details of *a posteriori* filtering can be found in [76].

From Figure 5.18, it is evident that with this approach we can estimate $s(n)$. For a comparison, Figure 5.18 also contains an estimate of $s(n)$ (denoted '$\tilde{s}(n)$') where the basis function in the SONF approach were assumed to be undamped (i.e., $\beta = 0 \rightarrow \phi_1(n) = cos(\omega_o n)$ & $\phi_2(n) = sin(\omega_o n)$). In practice, to simplify the problem, damped sinusoids are often modeled as undamped sinusoids. Clearly, the adaptive SONF with *a posteriori* filtering provides a method of adequately processing damped sinusoids with

unknown damping coefficients. Figure 5.19 shows the learning curves for β for a single trial and ensemble average (200 trials).

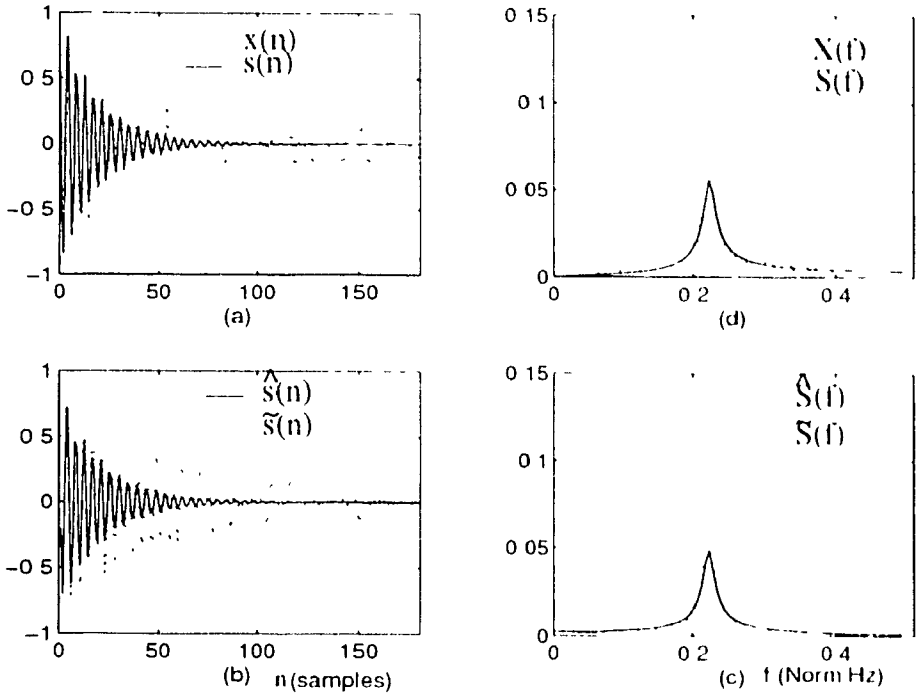


Figure 5.18: Results of Estimating $s(n)$ using the estimated damping coefficient $\beta(n)$ (a) Noisy input signal $x(n)$ and the true $s(n)$; (b) PSD of signal in (a); (c) Estimate of $s(n)$ using basis function with estimated damping coefficient, $\beta(n)$ and with $\beta = 0$; (d) PSD of signals in (c).

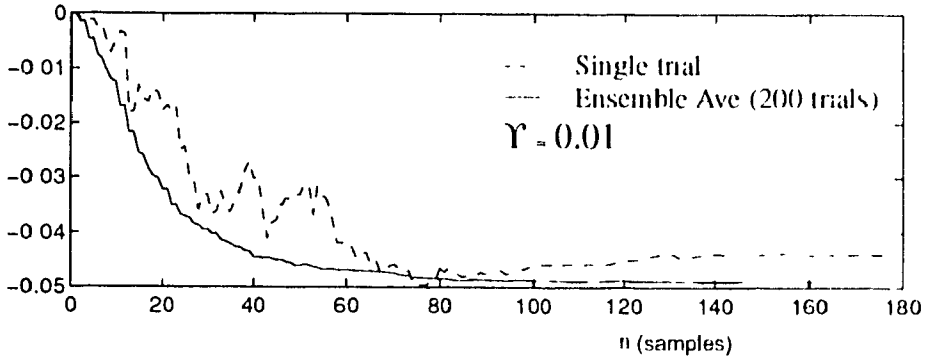


Figure 5.19: Learning curves of the damping coefficient, β.

Example 5.11 In this example, the input mixture is described as

$$x(n) = s(n) + r(n) + u(n) \tag{5.15}$$

where $s(n)$ is the desired or the message signal as in the previous example, $u(n)$ is zero-mean AWGN, and $r(n) = A_r e^{\beta n} \sin(\omega_r n + \theta_r)$ is the interference signal. The parameters β, -0.0202 ($e^{\beta} = 0.98$), $\omega_r = 0.2387$ Hz (normalized), and θ_r uniformly distributed over $[0, 2\pi]$, are chosen such that $r(n)$ is highly correlated with $s(n)$. The amplitudes A and A_r and noise variance are chosen such that the $SNR = 10$ dB and $SIR = -3$ dB (as defined in (5.3)). The step-size in the adaptive procedure is chosen to be $\gamma = 0.0009$.

For the same reasons as described in the previous example, we use the idea of *a posteriori* filtering. Figure 5.20 shows the results of extracting $s(n)$ from the input mixture. For comparison, we also extract $s(n)$ with the assumption that the desired signal is undamped, i.e., $\beta = 0$. This estimate is denoted with "$\hat{}$". Even in the presence of a strongly correlated interference we are able to use the Adaptive SONF to estimate $\hat{\beta}$, and hence to extract $s(n)$ from $x(n)$. Clearly from Figure 5.20, without the knowledge of β, we cannot estimate $s(n)$ with any degree of accuracy. Figure 5.21 show the single trial and ensemble average MSEs.

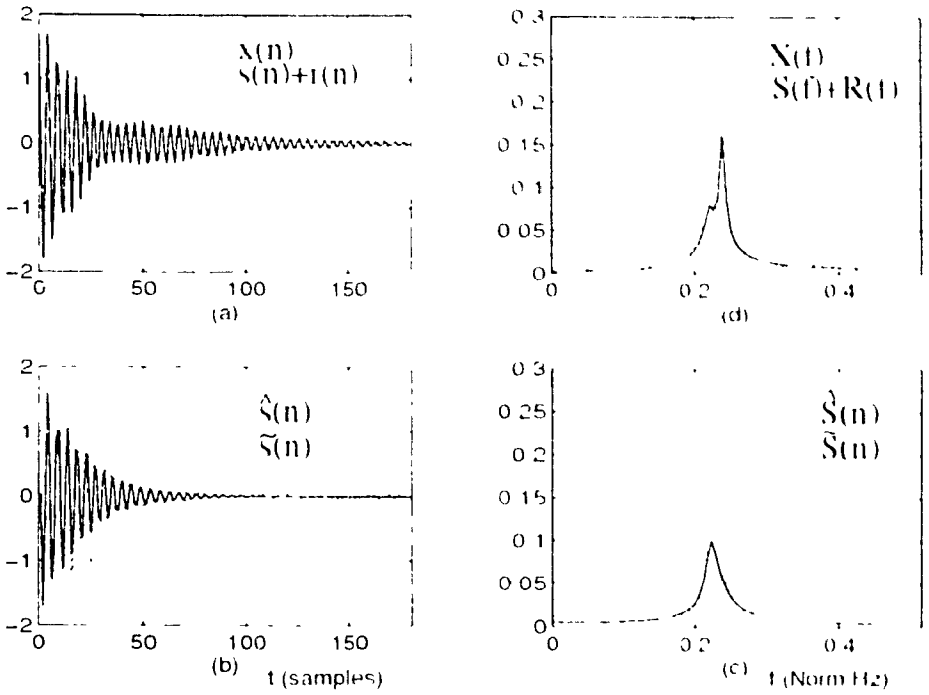


Figure 5.20: Results of Estimating $s(n)$ using the estimated damping coefficient $\beta(n)$ (a) Noisy input signal $x(n)$ and $s(n) + r(n)$; (b) PSD of signal in (a), (c) Estimate of $s(n)$ using basis function with estimated damping coefficient, $\beta(n)$ and with $\beta = 0$, (d) PSD of signals in (c).

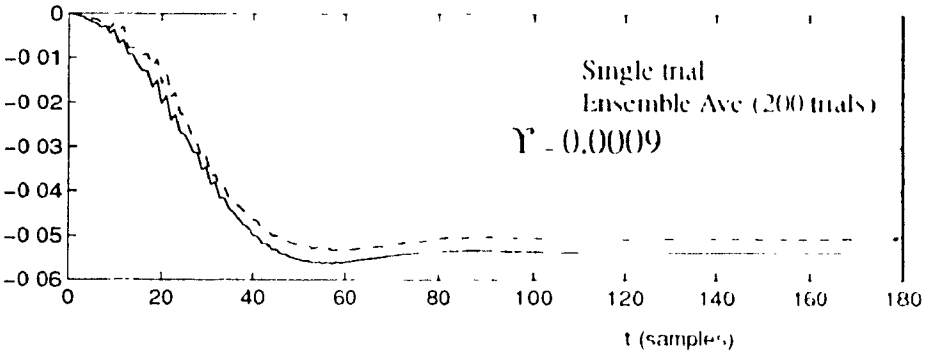


Figure 5.21: Learning curves of the damping coefficient, β

5.5 Conclusion

In this chapter, we have shown, via simulations, the effectiveness of the proposed approach of the SONF to process signals of short-duration. In comparison to the popular parametric ARMA model of CNF and the time-varying CNF (TV-CNF), it has been shown to be superior. In view of the fact that this approach is equivalent to the Kalman Filter (hence optimal in the mean-square sense), this is not an unexpected result.

Furthermore, since the SONF approach is time-varying, we have included simulation results to show how it can be used to process time-varying signals. We have also presented simulation results showing the use of the adaptive SONF, in situations where the basis functions representing the message signal are not completely known *a priori*. For damped sinusoids with unknown damping coefficients, the adaptive SONF is used to first estimate the damping coefficients and then the SONF is reapplied with this knowledge. Although not shown in this thesis, the adaptive SONF can always be constructed for any unknown parameters of the basis set representing the message signal.

Chapter 6

Auditory Evoked Brainstem Response (ABR)

Nonbehavioral hearing assessment has gained widespread acceptance in Audiometry. By nonbehavioral, we mean that a patient need not respond to whether a particular auditory stimulus was heard by him or not, thus, avoiding any patient bias or subjectivity. Evoked potentials or evoked responses (ERs) have gained an increasing acceptance as a nonbehavioral tool to measure acoustical sensitivity of the auditory path in the difficult to measure patients - infants, comatose, and handicapped [77]. From these responses, it is possible to determine various aspects of a patient's hearing capacity.

Auditory Evoked Brainstem Responses (ABRs) are obtained by applying an acoustical stimulus and then measuring the electrical activity of the brainstem by surface electrodes placed on the scalp. Measurement of this type impose one drawback, the measured response is embedded in the background EEG activity. Of course, we may use implanted microelectrodes in the brainstem to localize the interference by the background EEG activity, however, one can understand its lack of clinical feasibility for simple hearing loss test. Thus, a major engineering task is to reveal the ABR in the presence of the background EEG.

This chapter will explore the idea of ERs as a tool in auditory clinical environment. The first section will define "What are Evoked Responses?" The second section will describe the use of these evoked potentials in Audiometry. In particular, it will explore the waveform morphology of the auditory ERs, the method of acquisition and the parameters of interest from the response. Section 6.3 contains a brief summary of the current methods of evaluating the ABR waveforms. Section 6.4 states the problem at hand while Section 6.5 describes the procedure for processing the averaged ABRs. In Section 6 6, some illustrative results are presented indicating the ability of the proposed method to minimize the residual noise in the averaged ABRs. Finally, the last section contains some concluding remarks.

6.1 What are Evoked Responses?

Over the years many different tools for the analysis of the brain have emerged. One would think that its inner workings are well understood by now. Computer generated images of the brain anatomy fill number of medical journals, showing the results of NMR

and CT scans. None of these impressive techniques offer any idea of how the brain is functioning, its ability to think, or how the information is transported back and forth. From a psychiatric point of view, a useful tool would be a way to decipher human cognition or from neurological view how the brain controls various bodily functions [3].

There exists one primitive test, that does shed some light on how information is shuttled about the brain. By primitive we mean that since it was first introduced in 1950s by a physiologist named George Dawson, it hasn't changed much. Consider stimulating a patient's senses - a small electrical shock, a flash of light or a tone in the ear - which results in electrical responses being generated that eventually go to the brain. We are able to measure these responses at the brain by electrodes pasted to the scalp [3]. These types of responses are known as ERs. They are basically electroencephalograms (EEGs). But, unlike the ordinary EEGs (which show continuous electrical activity of the brain), these responses are elicited in a controlled manner. That is, ERs mark the brain's transient response to a particular stimulus, such as sound. In terms of audiological assessment, another way of looking at this would be to say: How does the brain react to hearing a sound? As the sound (the stimulus) travels from the ear, through the brainstem and to the higher brain, a series of pulses are generated along the way, causing scalp potential to vary with time. From this one can determine the reaction of a particular area along the auditory pathway [3].

If the scalp electrical recording is unusual - response appears later than expected, has a different shape than normal and so on - then a physician may conclude something is wrong. In this manner neurologists are using ERs to diagnose brain tumors, multiple sclerosis and hearing loss to name a few. Psychiatrists have hopes of using ERs one day to have insights into human cognition. Some researchers believe that someday, evoked potentials will help human-factor engineers design better cockpits, unbeatable lie detectors and even allow them to design equipment operated by a mere thought [3].

From the above discussion, it seems that the hopes of using ERs are great, but the obstacles in the development of this tool are also great. The problem with ERs is that they contain too much information. The electrodes not only pick up the response to the stimulus but also pickup the ongoing EEG - random firing of billions of nerve cells. Even the tiniest action of the body or the mind will result in some sort of electrical activity in

the brain. For example, electrodes detect activity from a blink of an eye, swallowing or even the 60 Hz noise from the nearby lightbulbs [3].

To extract the small ER (sometimes as low as nanovolts) in the midst of large background noise requires considerable amount of computing power. Even though the idea of ER was developed about forty years ago, it only found its way into medical laboratories as computers became inexpensive. Its full capacity can only be realized when more powerful software and processing algorithms are developed to separate various components of the EEG [3].

Most neurologists today measure ERs with the same processing tool used forty years ago. Dr. Dawson first applied the technique of signal averaging. He reasoned that by stimulating a patient's arm, he would be able to view the ER on an oscilloscope linked to the scalp electrodes, if he could find a way to enhance the 5 to 10 microvolt time-locked ER in the 50-100 microvolt EEG. He exposed his oscilloscope tracings to a photographic film. By stimulating the patient many times and superimposing the tracings on the film, the time-locked evoked response began to stand out amidst the random EEG. One major assumption in his reasoning was time invariance of the evoked response. That is, each stimulus will elicit a similar response in synchrony with the time of application of the stimulus. Today, instead of using a photographic plate a computer based averaging technique is used.

6.2 ERs in Audiometry

6.2.1 ABR Waveform Morphology

Dr. Dawson's reasoning in establishing the measuring technique of the ER was the repeatability of the response. That is, for every stimulus applied, the measured response will be similar. Irrespective of the patient's age, hearing sensitivity, or his waveform morphology, a repeated suprathreshold stimulus will produce a similar response [2, 77]. By suprathreshold, we mean stimulus level above the hearing threshold (i.e. lowest intensity level of the stimulus at which the brainstem response can be detected). Thus, repeated stimulus will produce a similar response.

Figure 6.1 depicts a typical suprathreshold waveform in time domain. The waveform

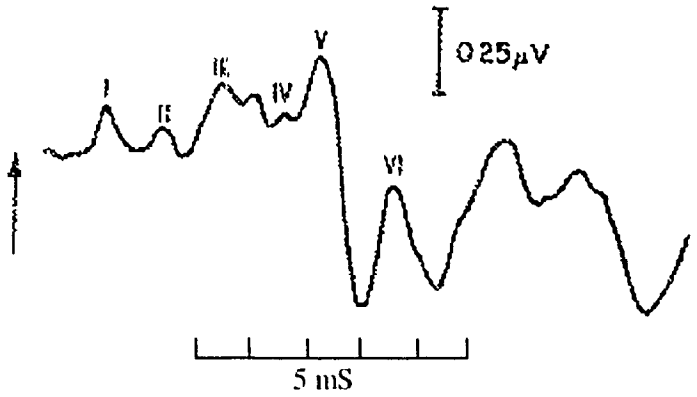


Figure 6.1: A typical ABR obtained at suprathreshold stimulus level. Adapted from [2]

of this figure is marked with an impulse at the start of the response. The impulse is only a representation of the stimulus to indicate the start of the response. Thus, a fraction of a second after the person hears a click (instead of tones, clicks are used as a stimulus for the ABRs), the ABR waveform is observed for the first 10 milliseconds. The response consists of a series of peaks and valleys. Figure 6.2 relates the seven components of the ABR to the different areas of the auditory tract as the electrical representation of the sound travels from the inner ear to the higher brain. The first component of the waveform is generated near the inner ear; components 2 through 5 arise somewhere in the brainstem and the location of generation of components 6 and 7 is still ambiguous. By the time the 10 milliseconds have elapsed the response has probably reached the cerebral cortex, the higher brain. In this region, the response is processed as cognition and is not useful as far as auditory evaluation is concerned. The locations of generation of these component responses in Figure 6.2 are only approximate as no one knows the exact point of generation. In all likelihood, there is more than one source for each component [3]. Figures 6.1 and 6.2 show a typical ABR waveform that is generated in normal healthy individuals. There exists a small variation in response from person to person, however, the general waveform shape is preserved. There have been attempts to categorize all possible waveforms into a finite set that can be used for template processing [78].

We have looked at the ABR in time domain up to now. We will now consider this response in frequency domain, as it will prove to be useful in the analysis or in the recovery of the signal from the noise. The spectral energy of the human EEG appears to

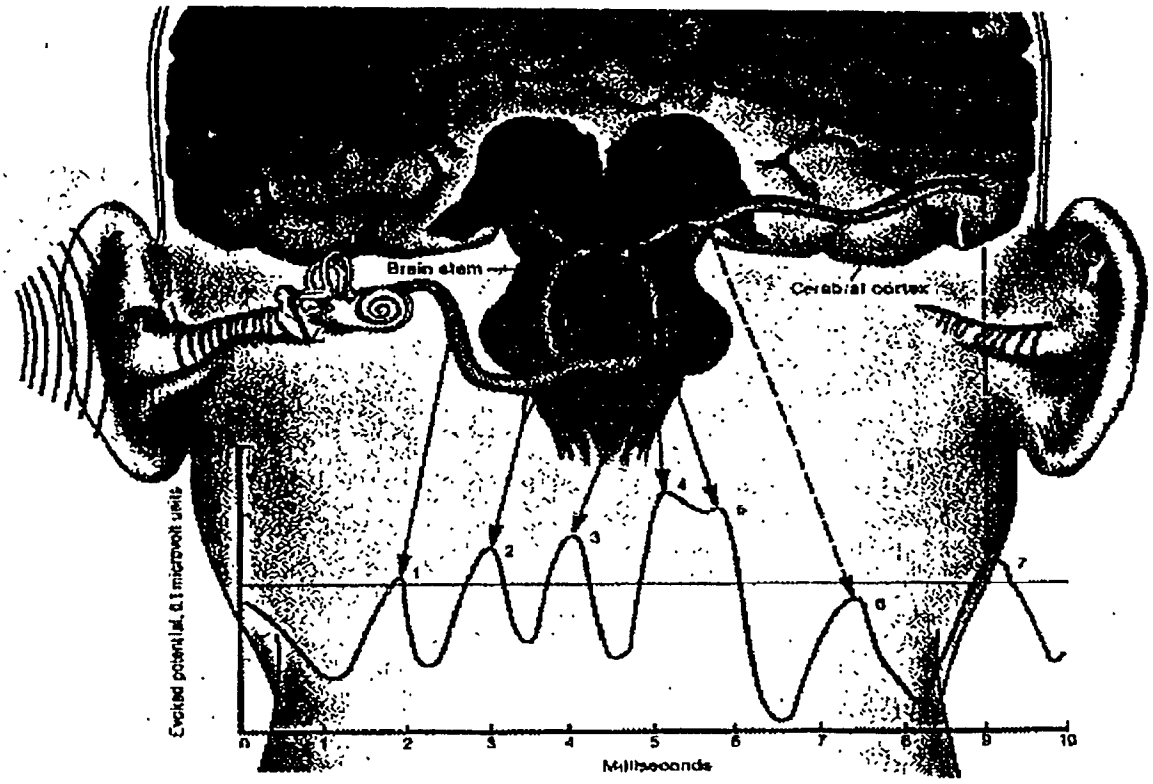


Figure 6.2: Location of generation of signal. Adapted from [3]

decrease with frequency up to about 2000 Hz and flattens above 2000 Hz. From the present studies available, spectral energy of ABR is limited to 1500 Hz [4, 79]. There are three main components: a low frequency component at around 100 Hz, (in the literature there appears to be some controversy about the existence of this component), a midfrequency component around 500 Hz, and a high frequency component around 900 Hz. Figure 6.3 shows a typical spectra of an ABR. Studies suggest that the midfrequency component is associated with component 5 of Figure 6.2, and the high frequency component is associated with the early components, 1 through 4. These two frequency components are probably the most important for hearing threshold and other audiometric diagnosis [4].

6.2.2 Measurement Techniques

ABR waveforms, unfortunately comprise about 1 percent [79] of the ongoing EEG activity and thus must be extracted by some signal processing means. The technique that was used forty years ago is still in practice today - the idea of signal averaging to eliminate noise.

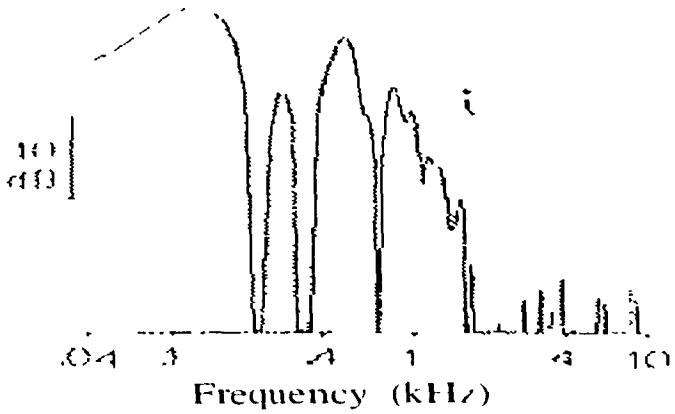


Figure 6.3: Spectra of ABR. Adapted from [1]

Consider a digitized signal

$$x_j(k) = s(k) + n_j(k) \tag{6.1}$$

(k refers to discrete time and the subscript j refers to jth run in the averaging process) where $s(k)$ is the evoked response and $n(k)$ is the background noise. Now, if the noise is assumed to be white (there is no correlation in individual samples) and is uncorrelated with $s(k)$, then we can enhance the signal $s(k)$ by applying signal averaging technique. Each response x_j will have a noise component n_j that is uncorrelated from response to response (results of whiteness of the noise). By using this random nature, it is possible to sum a large number of responses to reduce the effect of noise [80]. In the equation

$$\hat{s}(k) = \lim_{M\to\infty} \frac{1}{M} \sum_{j=1}^{M} x_j(k) = s(k) + \frac{1}{M} \sum_{j=1}^{M} n_j(k) \tag{6.2}$$

if M approaches a large number then the effect of the noise can be minimized, revealing a clean ABR component. In (6.2), M is the number of runs, $s(k)$ is the ABR, $n_j(k)$ is the noise component in the jth run and $\hat{s}(k)$ is the estimated ABR. If the ABR component of the signal x is not similar from run to run, then it is obvious that some of it will also be canceled. For a further discussion of the minimization of noise in ERs, see [80, 81].

There are many different ways to record these responses, one such system is shown in Figure 6.4. Standard EEG disk electrodes are attached to the vertex, the mastoid and, the forehead which is used as a ground reference. The preamplifier can have a variable gain which is usually around 100 within the cutoff frequencies of about 100 and 3000 Hz.

The output of this preamplifier is connected to an amplifier of a gain around 1000, which is also adjustable. Finally a microcomputer based data acquisition system is used to average and store the waveforms. The applied auditory stimulus is a very small pulse with its intensity adjusted by a decade counter. The reason for using a pulse instead of a tone is that a pulse covers the full frequency spectrum. Thus, it is capable of exciting the neurons in the auditory tract at all frequencies, not just a single tone. Through an earphone, the click or the pulse stimulus is monaurally delivered to a patient relaxed in a supine (on their back) position in a sound-attenuated chamber. A typical ABR testing procedure for hearing thresholds will involve a collection of a set of responses - one for each stimulus level for each ear [79, 82].

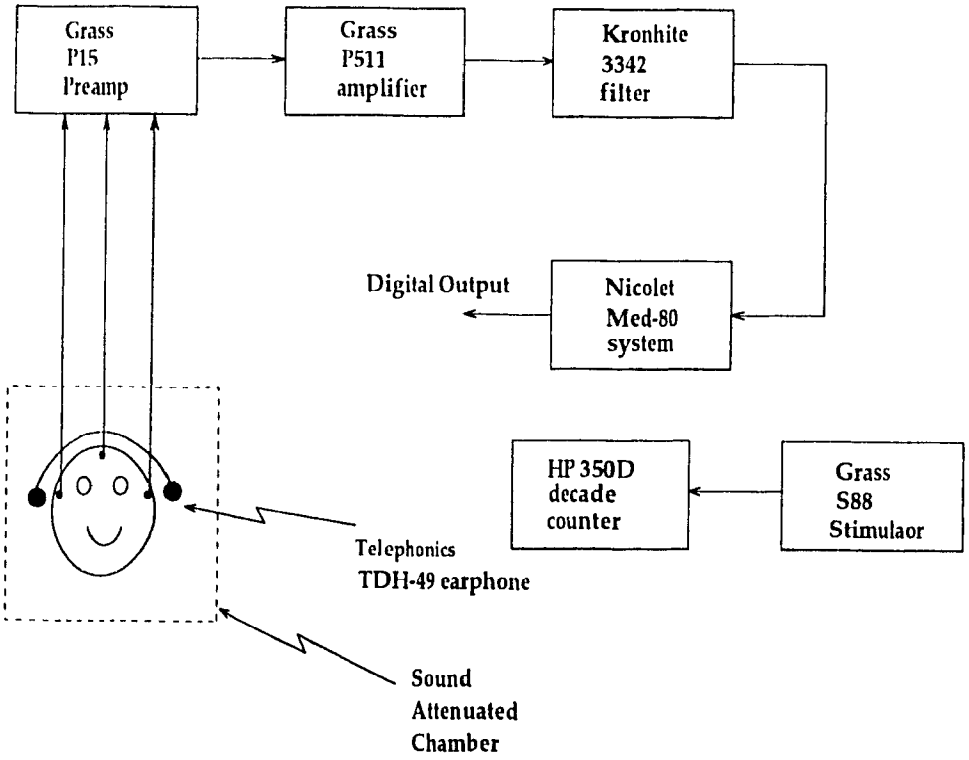


Figure 6.4: ABR acquisition setup.

The measurements are obtained in an ipsilateral configuration (stimulus is applied to the ear that is being tested) , however, contralateral (stimulus is applied to the ear opposite to the one being tested) measurements are taken in certain cases. In a general

procedure, the examiner begins by obtaining a response with a high (suprathreshold) stimulus level and then in a predetermined step size decreases the stimulus level to obtain a set of responses for each ear. He also obtains a response with no stimulus that serves as a control. Each ear is tested independently.

In the determination of the hearing "threshold" or the lowest intensity level at which a response is detected, it is necessary to collect averaged responses at various intensity levels and then determine whether an ABR is present or only background EEG exists. Thus, for this purpose, the only parameter of interest is the amplitude of the ABR. That is, it is only required to distinguish the presence or absence of the ABR, rather than its shape. At intensities much higher than an individual's threshold, ABRs are easily detected upon visual inspection, however, as the intensity approaches an individual's threshold, the response is marred by the background EEG activity and it becomes a difficult task for the clinician to evaluate these waveforms. Figure 6.5 shows a set of ABR waveforms for the left ear of one subject. Evaluation of this set yielded a threshold of 50 dB. Notice that the waveforms at much higher stimulus levels than the threshold are clean and easily detected, however, at levels below and close to the threshold, it is difficult to determine the presence or absence of the ABR [1,10,11].

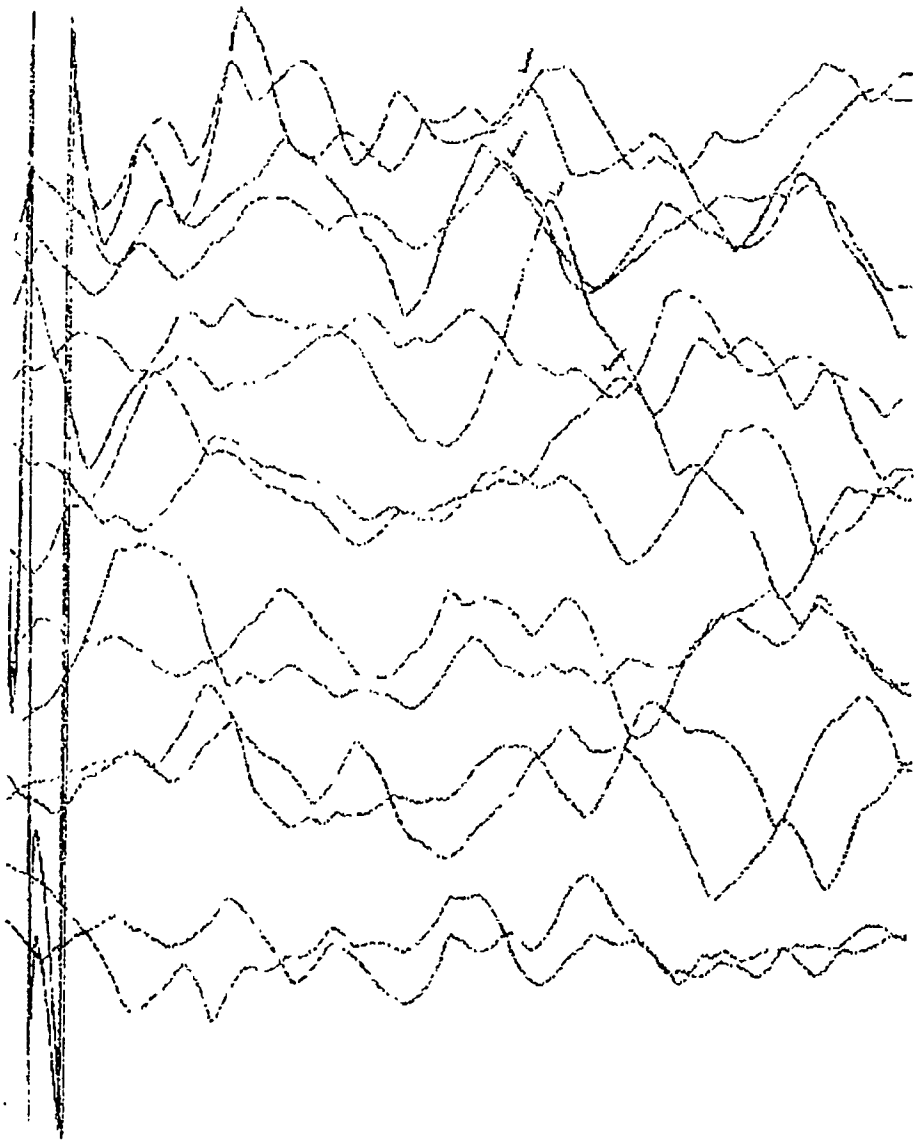


Figure 6.5: An averaged set of ABR for one ear [5].

6.3 Present Methods of Evaluating the Acquired ABRs

Whether an ABR is judged as being present or absent depends on the tester's experience in evaluating the waveforms. These judgments of the waveforms can become subjective since the tester may have some prior knowledge of the patient history or he may simply judge the waveforms differently on different days. When the ABRs are particularly noisy, the tester's bias becomes increasingly important and may lead to erroneous evaluations. Another problem that exists in this kind of visual evaluation is intertester reliability. Different testers may adopt different criterion for accepting a waveform as containing an ABR. Thus, evaluation of the ABRs for hearing threshold is not a *wholly* objective procedure [83].

To eliminate the subjectivity in the evaluation process, various ideas have been proposed in the literature. The goal has been to develop a mathematical tool to detect the presence of the ABR. Weber and Fletcher [77] have proposed an algorithm based on correlation between pairs of waves obtained with and without acoustic stimulus. Wong and Bickford [84] compared the power in the ABR waveforms with and without acoustic stimulus. Salamon [85] applied a nonparametric analysis of the variance test. All the above algorithms were found to have the same statistical validity as the visual detection method, however no extensive studies have been done to dismiss or accept the usefulness of these methods. By visual detection, we mean that a clinician scores the presence/absence of the ABR by looking at the averaged waveform. Recently, Arnold [82] compared visual detection against (1) correlation [85], (2) variance ratio [84], and (3) pre and poststimulus differences [86]. The results indicate that the visual detection was the most sensitive, but it was not more statistically sensitive than the correlation method.

6.4 Problem Statement

In practice it usually requires about 2000 sweeps (individual responses) for each stimulus intensity level to acquire an averaged ABR. A typical ABR recording session lasts from 45 60 minutes. Due to the length of time required to acquire these waveforms, patient starts becoming restless which results in severely degraded response. Additionally, in practice,

the background EEG or noise is neither white nor stationary [79], thus allowing for some noise to be also enhanced. If the number of responses in the averaging is reduced, as is done to reduce the length of time required to collect a set of ABRs, the noise is not sufficiently suppressed. Application of conventional techniques to reduce this residual noise is limited due to the short duration nature [79] of the ABR waveforms.

A method to reduce the residual noise component in the short-duration averaged ABR is required. In the next section, we propose a method that can be used to process the averaged ABRs to further reduce the noise. The proposed method is a two step approach that combines the Constrained Adaptive Notch Filters (CANFs) [26] and Statistically Optimal Null Filters (SONFs) of Chapter 4.

6.5 Method

In literature there is ambiguity about the existence of the lower component (100 Hz component). Thus, the two important frequency components centered around 500 and 900 Hz are of primary interest in audiology [4]. Moreover, in literature, these components have been modeled as damped sinusoids [70, 87] due to the transient like nature of the ABRs, i.e ,

$$s(k) = s_1(k) + s_2(k) \tag{6.3}$$

$$= A_1 e^{\beta_1 k} cos(f_1 k + \theta_1) + A_2 e^{\beta_2 k} cos(f_2 k + \theta_2) \tag{6.4}$$

where A_i's, β_i's, θ_i's and f_i's are respectively the amplitudes, damping factors, random initial phases and the frequencies of the different NB component of the ABR. Equation (6.4) can be written in terms of the basis set $\{\phi_i(t), i = 1, 2, 3, 4\}$ as

$$s(t) = \sum_{i=1}^{4} v_i \phi_i(t) \tag{6.5}$$

where c_i's are random variables and

$$\phi_1(t) = e^{-\beta_1 t}\cos(f_1 t)$$

$$\phi_2(t) = e^{-\beta_1 t}\sin(f_1 t)$$

$$\phi_3(t) = e^{-\beta_2 t}\cos(f_2 t)$$

$$\phi_4(t) = e^{-\beta_2 t}\sin(f_2 t)$$

are the basis functions. The basis functions $\phi_1(k)$ and $\phi_2(k)$ form a pair since they model one component of the ABR. Similarly, $\phi_3(k)$ and $\phi_4(k)$ form another pair.

It is assumed that the parameters $\{f_1, \beta_1, f_2, \beta_2\}$, and hence the basis functions are *a priori* available. We can therefore use the signal estimation counterpart of the SONF (as shown in Figure 4.11) to estimate each term of (6.5). Since this is a linear time varying filter, we can use the idea of superposition to estimate all the four components of (6.5). This is described by the signal estimation counterpart of the NC SONF, shown in Figure 4.13 with $N = 4$. Note that with this strategy, we may also consider the 100 Hz component of the ABR, if the need or justification arises.

An important criterion in using SONFs is that the basis functions must be orthogonal for every time instant $t \in [0, T]$. This can be achieved using a sliding Gram Schmidt orthogonalization procedure as described in Chapter 4. Additionally, since there is no *a priori* knowledge of the SNR of each component of (6.5), we must use the suboptimal $\lambda_i(t)$, i.e. $\lambda_i(t) = \phi_i(t)/\|\phi_i(t)\|_t^2, \quad i = 1, 2, 3, 4$. This, of course, will provide suboptimal estimation. Alternatively, since the measured ABRs are in a digital form, we can avoid the computationally expensive sliding GS orthogonalization procedure and use the recursive SONF as outlined in Chapter 4 to obtain a *globally optimal estimate*. Given that we do not know the SNR, we use the identity matrix as $P(0)$.

Parameter Estimation

Thus far, we have tacitly assumed that the frequencies f_i's and the damping factors β_i's are *a priori* available. Although we have some idea about the frequencies, this is generally not the case for the damping factors, β_i's. These parameters must be first estimated, with

which the basis functions can be generated. It is only then, that we can use the SONFs of Chapter 4. We now describe how these parameters can be determined.

In determining the hearing "threshold" or the lowest stimulus intensity level at which an ABR can be discernible, it is necessary to collect a set of averaged responses at various intensity levels. At levels much higher than an individual's threshold, ABRs are easily detected. The averaging technique yields a well defined waveform. We use these suprathreshold waveforms, to estimate the aforementioned parameters – center frequencies and the respective damping factors. Due to the assumed replicability of the ABR, we can use these estimated parameters for *all* ABRs obtained at lower intensity levels for the same patient or subject.

The parameters can be estimated in two parts. First the frequencies f_1 and f_2 are estimated using the CANFs as described in [76, 88]. The estimates of the frequencies $\hat{f}_1$ and $\hat{f}_2$ (hat denotes estimate) are then used to find the damping factors β_1 and β_2. The damping factor for each NB component of the ABR is estimated separately. Figure 6.6 describes the basic idea in estimating β_1. Notice that to estimate each β_i, a second order signal enhancement counterpart of the NC-SONF is required – dashed box in Figure 6.6 as a basic building block in the adaptive configuration. A similar section in cascade/parallel can be used to estimate β_2. Steepest-descent algorithm is used to minimize the output error in the estimate of the ith component of the ABR with respect to β_i. At each instant of time the current estimate of $\hat{\beta}_i$ is used to generate new basis functions, $\hat{\phi}(l)$'s. Section 4.7 describes in detail the adaptive algorithm for estimating the damping factors, β_i's.

6.6 Results

In this section, we will illustrate the effectiveness of the proposed approach by presenting some example results of filtering averaged ABRs to reduce the residual noise. We use the term averaged ABR to mean enhanced or preprocessed ABR by the averaging technique described earlier. Sets of ABR with various stimulus intensity levels were collected using an acquisition system described in Figure 6.4 for hearing threshold assessment [5]. Each set consisted of several ABRs at different intensity levels and for each stimulus intensity

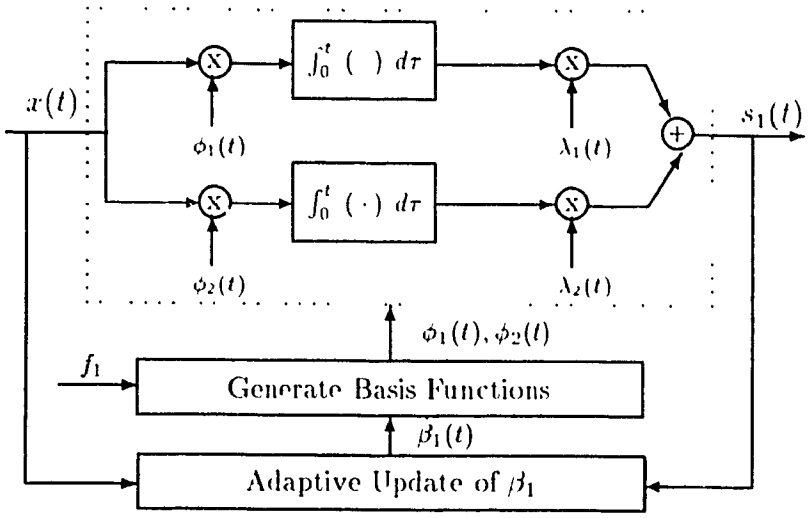


Figure 6.6: Adaptive estimation of the damping factor β_1 using the NC-SONF as a basic block.

level two averaged ABRs of 10.24 ms duration were obtained. The sampling rate was set to 25 kHz and 1024 such responses were averaged to form each averaged ABR. Estimates of f_1, f_2, β_1 and β_2 were obtained for the **two averaged ABRs** elicited at the highest stimulus intensity levels in each set (where the ABRs were clearly evident). Before the parameters were estimated, each of the two averaged ABRs were filtered with a fourth order bandpass Butterworth filter with cutoff frequencies of 300 and 1500 Hz to remove the stimulus artifact. Before estimating the frequency and the damping factor, both averaged ABRs were normalized to unity. The average of the two estimates for each parameter was then used to generate the required basis functions for the estimation of the ABR waveforms via the recursive NC-SONF. Figure 6.7 outlines the general procedure in the filtering of the ABRs.

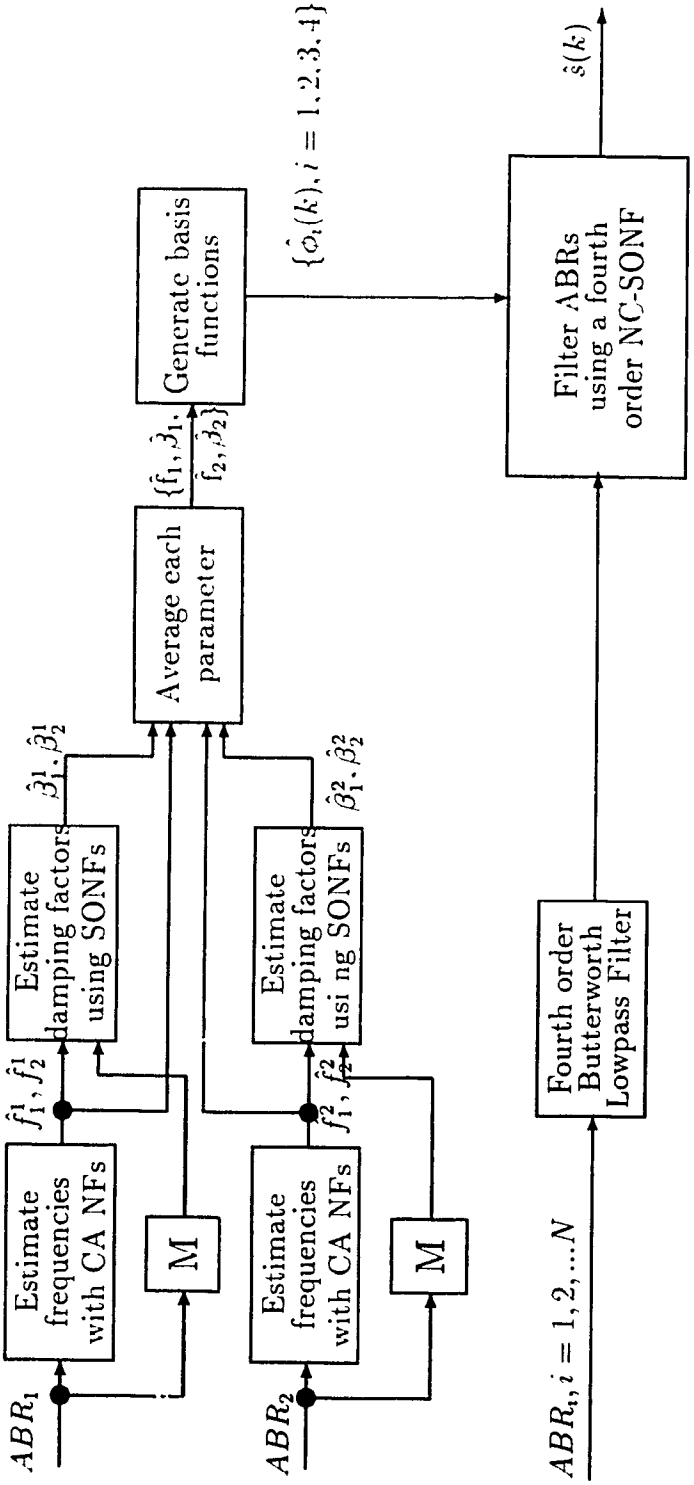


Figure 6.7: Configuration to filter the averaged ABRs in threshold testing.

Example 6.1 Figure 6.8 shows the result of one ABR in a typical set [89]. The stimulus intensity level for this ABR was 10 dB lower than the ones used for parameter estimation. Clearly the proposed algorithm applied to the averaged ABR yields a waveform where the ABR is well defined and allows for more reliable detection.

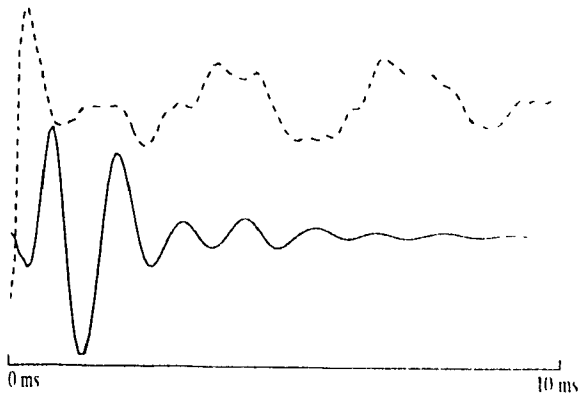


Figure 6.8: Results of Proposed Methods: filtered ABR; - ABR as acquired.

Example 6.2 Figure 6.9 shows a set of ABRs[1] (i.e. as acquired). The numbers in parenthesis on the right indicates the stimulus intensity in dB and the solid and dashed lines indicate two different averaged waveforms acquired for the same intensity level. The last pair in the set is the silent control, i.e. response obtained without applying a stimulus.

Examining the original (averaged) waveforms of Figure 6.9, it was ascertained by an audiologist that they showed no response. It was concluded that this is severely hearing impaired subject, however, it may still be possible for the high levels of the physical activity (motor activity) to influence the outcome. Table 6.1 shows the results of the correlation method. The second column represents the cross-correlation of the two ABRs obtained at the same stimulus intensity level while the third column represents the cross-correlation of the silent control. Column four shows the measure used to decide the presence/absence of the ABRs. It is based on the four combinations of the cross correlation between the two acquired ABRs at the same intensity level and the two silent control waveforms. With the measure 4/4 represents a clear presence and 0/4 represents the absence of the ABR. From these results, it is difficult to determine a threshold level. The criterion for the detection

[1]The waveforms have been collected at the Royal Victoria Hospital under the direction of Dr. H. J. Hecki [5]. The author had the opportunity to spend several months collecting new data and transferring some existing data for this project.

of ABR by the correlation method [85] is not met consistently enough to be able to make a threshold decision. Figure 6.10 contains the ABRs enhanced via the SONF procedure of Figure 6.7. Evaluating these waveforms, it can be roughly seen that there exists a response for stimulus intensity as low as 55 dB, however, it is likely that a trained clinician may accept the waveforms as containing a response for even lower intensity levels. In this latter set, various components of each ABR appear to be defined more clearly than the simply averaged waveforms of Figure 6.9. Thus, other parameters in ABR audiometry (such as latency) can be much more readily measured.

Hearing Sensation Level	Stimulus CC	Silent Control CC	Response
91	.12	-.23	1/4
85	.15	-.23	1/4
75	.54	-.23	4/4
65	.43	-.23	2/4
55	.13	-.23	0/4
45	.36	-.23	4/4
35	.41	-.23	3/4

Table 6.1: Correlation Method Analysis of ABRs of Figure 6.9.

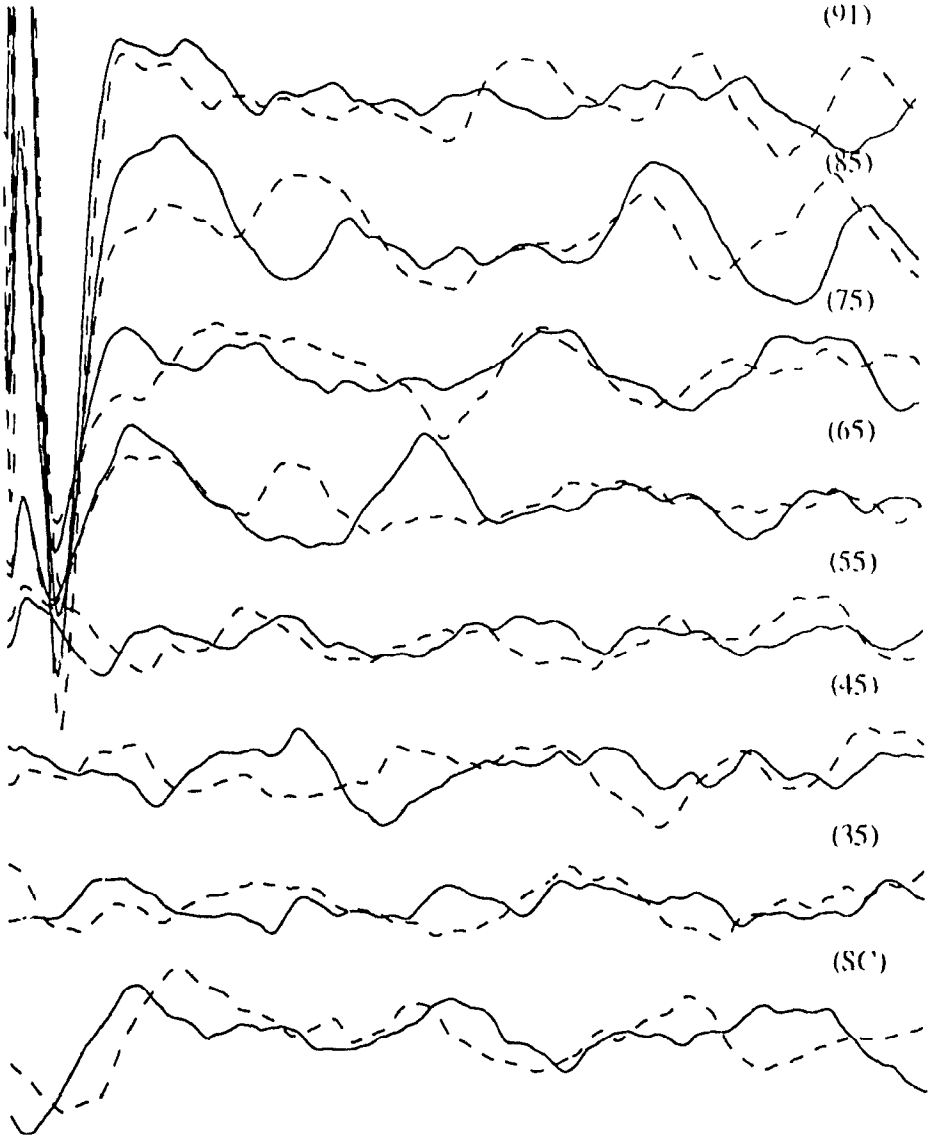


Figure 6.9: ABRs collected from a left ear of a patient [5].

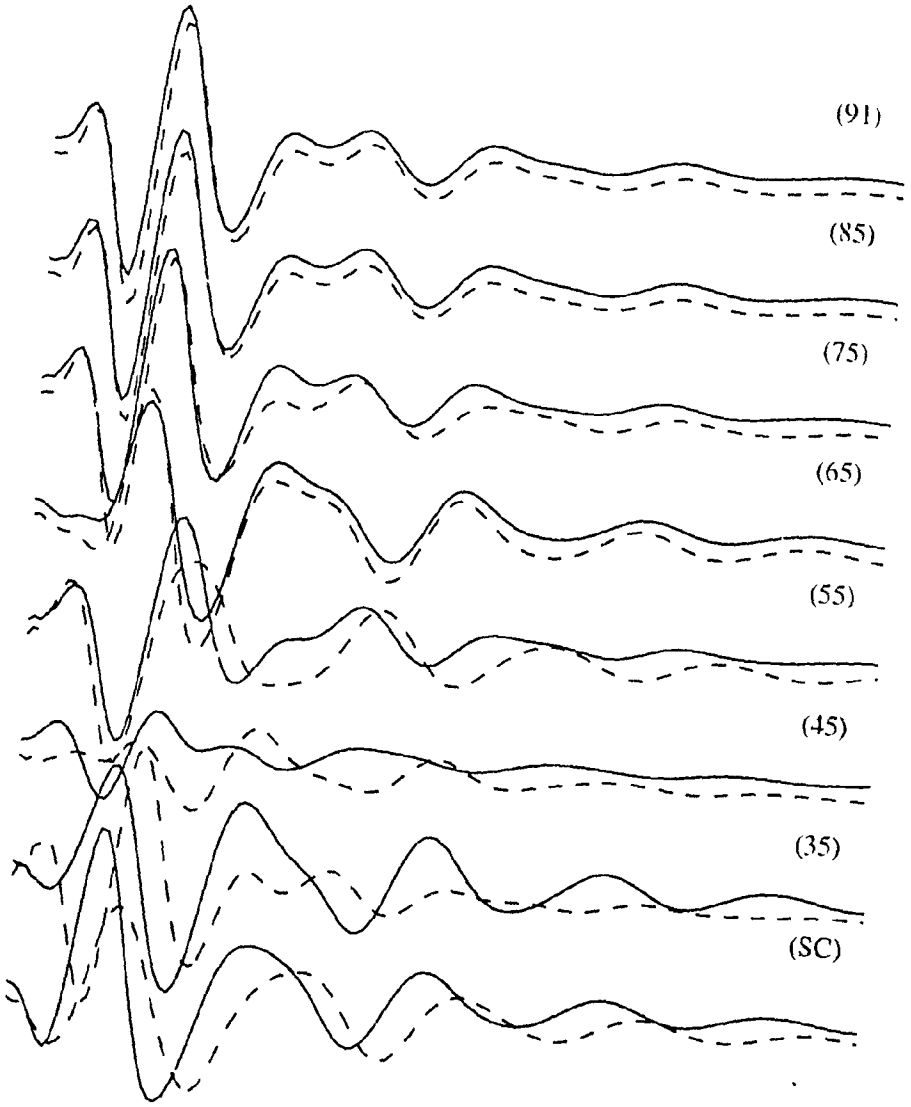


Figure 6.10: The set in Figure 6.9 filtered via the recursive SONF.

Example 6.3 Similar to the last example, Figure 6.11 contain the recording of another set of ABRs at various stimulus intensities and Figure 6.12 contains ABRs processed by the proposed strategy. In this situation, based on the visual inspection of the waveforms of Figure 6.11 by an audiologist, profound hearing loss of this ear is concluded. Correlation method yielded a threshold of 85 dB (see Table 6.2), inconsistent with that of the visual inspection. Examination of the waveforms in Figure 6.12, yielded a threshold at the lowest intensity level (75 dB) The individual peaks in Figure 6.12 are much more clearly defined. Should the test have continued for lower intensities, it is possible (even likely) that the hearing threshold is much lower.

Hearing Sensation Level	Stimulus CC	Silent Control CC	Response
91	.71	.37	1/1
91	.85	.37	1/1
91	.32	.37	3/1
85	.17	.37	2/1
85	.61	.37	1/1
75	.33	.37	0/1

Table 6.2: Correlation Method Analysis of ABRs of Figure 6.11.

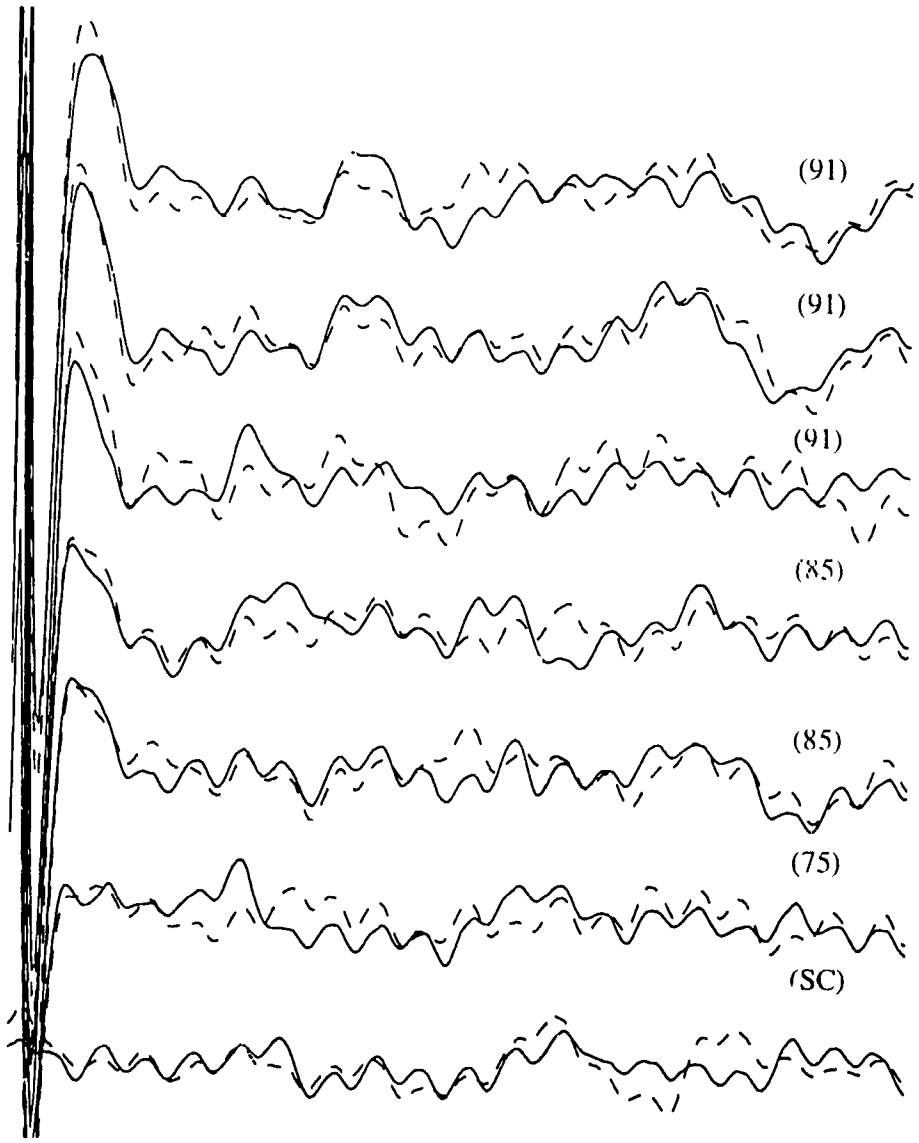


Figure 6.11: ABRs collected from a left ear of a patient [5].

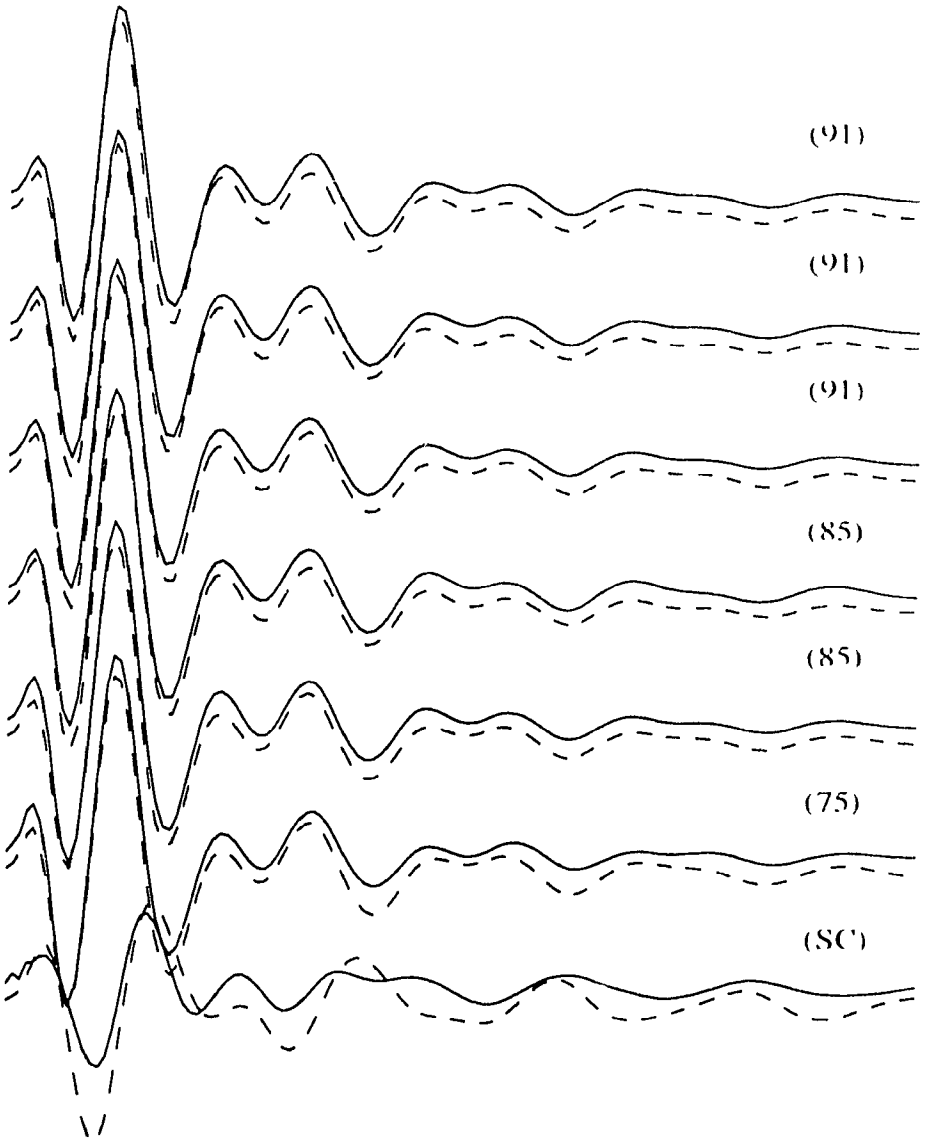


Figure 6.12: The set in Figure 6.11 filtered via the recursive SONF.

6.7 Conclusion

As the number of trials in the averaging process to enhance the ABRs is reduced, the noise level can become significantly high. In an attempt to minimize this residual noise, we have proposed a two-step algorithm that uses a sum of damped sinusoids model of the ABR. In the first step, the sinusoid frequencies and the damping factors are estimated, while in the second step these parameters are used to enhance the averaged ABRs by the recursive SONF in an *a posteriori* filtering [76] scheme.

Using this approach, results of filtering two sets of ABRs from two independent ears have been presented. Both the visual inspection and the correlation method failed for one case, and for the second case they showed inconsistencies. Results of filtering these waveforms with our approach provide clearly defined ABRs where none were visually apparent in the original unfiltered waveforms. Upon visual inspection of filtered ABRs in Examples 5.2 and 5.3, it is likely that a clinician would change the original conclusion of severe deafness. Furthermore, the averaged ABRs filtered by our approach can be used in the correlation method to yield automatic detection.

The results presented in this Chapter indicate that we may use the SONF based method to minimize the residual noise in the averaged ABRs leading to the possibility of a more reliable detection procedure, perhaps even a completely objective procedure. Moreover, with the ability to minimize this residual noise, it may be possible to reduce the number of trials required in the averaging process for each ABR. Given these favorable results, a statistical study of the presented method is essential before it can be accepted or rejected.

Chapter 7

Conclusion

The problem of processing transient signals, whether for suppression, enhancement or classification, generally presents a difficult task. The difficulties arise due to the limited duration of the data and are further compounded by low SNR, low SIR, limited *a priori* information and so forth. In this thesis, we have addressed the problem of processing such signals in the presence of noise given their existence and in some applications, their TOA. For such signals steady-state approaches clearly are not applicable. Thus, in general, we have addressed the problem of estimating short duration signals in the presence of noise, irrespective of whether they are of decaying type or not.

The methods employed in the processing of signals are ultimately determined by their mathematical description or representation. In this thesis, we have considered two forms of signal modeling within the Structural Signal Processing (SSP) framework to generate null filters for processing short duration signals. The first is the parametric model that leads to the Almost-Symmetric ARMA (AS-ARMA) form of null filtering. In the second, the more classic series expansion model is used to generate the new nonparametric Statistically-Optimal Null Fi (SONFs).

In the first part of the thesis, we have concentrated on the parametric form of null filtering. For practically useful CNFs (in terms of filter transient duration), the distortions in the frequency response reduce the allpass nature of the filter. In this thesis, it has been shown that such distortions can be corrected by increasing the order of the CNFs based on strategic pole/zero pair placement. Three closed-form methods of finding such strategic pole/zero pair locations have been derived. These filters have been termed Modified Notch Filters (MNFs). Extensive simulation results show the improvements in the MNFs over CNFs. The improved allpass performance of the MNFs is then traded to obtain shorter filter transient durations for the processing of short duration signals.

The degree to which AS-ARMA null filters are transparent to signals not belonging to their null space is determined by the symmetry factor, α. To quantify the measure of transparency of these filters, a new criterion is developed. An expression in terms of filter coefficients is derived.

In the second part of the thesis, we exploited the series expansion representation of the considered signal. The knowledge of the set of basis functions in this expansion has been used to develop the new nonparametric SONFs.

It is well known that MFs increase the output SNR in a very short time, thus, we have used a special Instantaneous MF (IMF) as a basic building block in our approach. By combining the maximum output SNR and the LS criteria, the SONF has been developed. Several different forms of the SONFs are presented and the relative advantages/disadvantages of each are discussed in the thesis. Detailed analysis reveals the SONF to be equivalent to the well known Kalman Filter. As such, it offers a new and unique way of implementing the Kalman Filter. One of the key advantages of the SONF is realized in the simplicity of its implementation. This simplicity contrasts with the more involved Kalman Filter, yet the end result is the same.

To further motivate the problem of enhancement/suppression of short duration transient-like signals, we have considered the problem of Auditory evoked Brainstem Response (ABR) audiometry. It has been shown that by using a combination of adaptive and recursive SONFs, it is possible to minimize the residual noise in the averaged ABRs. Real data examples have been used to illustrate the effectiveness of such an approach and to warrant further statistical evaluation of the method.

Main contributions

Contributions of this thesis are centered around the theme of short duration signals. The following briefly highlights the main contributions of the thesis.

- In order to describe the intrinsic properties of parametric Almost Symmetrical ARMA based Null Filters, we have developed a new performance criterion, the *transparency imperfection coefficient*, Γ. It describes the quality of the passband in terms of the filter coefficient.

- It is shown that by increasing the order of the CNFs by strategically placing pole/zero pairs, the frequency response can be improved.

- Improvements in the frequency response of the null filters can be used to reduce the duration of the filter transients and enable them to be used for the analysis of short duration signals.

- Three closed form methods of evaluating the additional pole/zero pairs locations have been developed. One is optimal while the other two are suboptimal.
- A new nonparametric SONF in continuous-time is presented for the processing of short duration or transient signals. It is also shown that such an approach is only locally optimal.
- Globally optimal SONF is presented
- To deal with the implementation issues of the SONFs, a discrete-time recursive version of the SONF is developed.
- For the case where the basis functions used in representing the series expansion of the signal under consideration are not completely *a priori* known, an adaptive version of the SONF is presented.
- Since the SONF is LTV and optimal under the MSE criterion, it is believed that it is closely related to the Kalman Filter. This is shown to be true analytically for the first order case. For higher order SONFs, simulations are used to justify this claim.
- Extensive simulation results show drastic improvement in the performance of the SONF over the parametric null filtering methods for the short data signals.
- Simulations are used to show that SONF can be adequately used for the processing of signals with time-varying parameters.
- Application of the SONF in the separation of highly correlated signals of given structure is presented.
- SONF based approach is applied to filter the averaged ABR waveforms using a sum of exponentially damped sinusoids model. Several examples of filtering these waveforms indicate the usefulness of this approach in the application of nonbehavioral hearing loss assessment of humans.

Future Research Topics

Interesting topics for future research include:

- Application of MNF as an adaptive notch filter to estimate the unknown NB component frequencies may be explored. We expect the resulting frequency estimates to be of lower bias and variance as compared to the CANFs of [26] due to the improved frequency response of the MNFs.

- The removal of the constraint on the pole locations of the compensation section of the MNF and determination of optimal locations using nonlinear optimization techniques. This may further improve the frequency response of the MNFs.

- For processing of time-varying signals of unknown waveform shape the ASTV ARMA have been recently developed [32]. Many different structures of the ASTV ARMA have been proposed. Since the SONF approach is equivalent to the Kalman Filter (thus, valid for time-varying signals) and given the simulation results in Chapter 5, it will prove interesting to modify the different structures of ASTV ARMA presented in [32] using the SONF as a key building block. Simulation results indicated that the SONF is superior to the ASTV-ARMA null filter for short duration signals, we expect the SONF based methods to be superior for time-varying signals as well.

- The limits of integration in the IMF are from 0 to t. As $t \to \infty$, the norm $c(t)$ $\int_0^t \phi^2(\tau)d\tau$ will grow to a very large value. Hence, new ways must be devised to limit this growth. One possibility is to use a sliding time window. That is, use limits from $t = t_o$ to $t = t_f$, where the interval $t_f - t_o = T$ is constant at each time instant. Alternatively, one may employ a forgetting factor as is done in the RLS algorithm. This procedure is essential if the SONF is to be used for a signal over long observation intervals.

- We have applied the CANFs and the different SONFs in removing the residual noise in the averaged ABRs. Results presented are only examples showing the viability of this approach. A rigorous statistical study must be done to adequately evaluate the usefulness of this approach.

- Application of the approach of Chapter 6 to single trial ABRs or to waveforms obtained with reduced number of trials in the averaging procedure.

- In this thesis, we have modeled the ABRs as damped sinusoids as is done in literature. Although this provides us with simple basis functions for the SONF approach, it would prove interesting to try different basis functions. As an initial suggestion, **sinc** or some other mother wavelet family of basis functions may be tried. Comparisons of results based on different basis functions in the SONF approach may reveal or enhance different properties of the ABR.

- It may be useful to estimate the ABRs using the traditional ALE using the LS solution rather than the LMS algorithm. Recent improvements/suggestions such as the choice of model order, the proper choice of delay for decorrelation (Δ), and data reusing may be incorporated. The results can be compared to the SONF approach.

Bibliography

[1] R. Agarwal, E. I. Plotkin, and M. N. S. Swamy, "Statistically optimal null filter for processing short record length signals," in *Proceedings of ISCAS '95*, (Seattle, Washington), April 1995.

[2] K. Hecox and R. Galambos, "Brainstem auditory evoked responses in human infants and adults," *Arch. Otolarynology*, vol. 99, Jan. 1974.

[3] C. Truxal, "Watching the brain at work," *IEEE Spectrum*, Mar. 1983.

[4] J. R. Boston, "Spectra of auditory brainstem responses and spontaneous EEG," *IEEE Trans. Biomedical Engineering*, vol. BME-28, Apr. 1981.

[5] *ABR Data*. Royal Victoria Hospital, Department of Otolaryngology. This data was obtained with the help of a clinician under the direction of Dr. H. J. Hecki

[6] M. Frish and H. Messer, "Detection of a known transient signal of unknown scaling and arrival time," *IEEE Trans. Signal Processing*, vol. 12, July 1994.

[7] B. Porat and B. Friedlander, "Performance analysis of a class of transient detection algorithms – a unified framework," *IEEE Trans. Signal Processing*, vol. 40, Oct. 1992.

[8] B. Friedlander and B. Porat, "Detection of transient signals by the gabor representation," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-37, pp. 169–179, Feb. 1989.

[9] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.

[10] I. Daubechies, *The wavelet transform: A method for time-frequency localization. Advances in Spectrum Analysis and Array Processing*, vol. 1. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[11] J. A. Cadzow and M.-M. Wu, "Analysis of transient data in noise," *IEE Proceedings, Part F*, vol. vol, Dec. 1987.

[12] S. L. Marple Jr., *Modern Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[13] T. A. C. M. Classen and W. F. G. Mechlenbrauker, "The wigner distribution - a tool for time-frequency signal analysis: Part I: Continuous-time signals," *Phillips J. Res.*, vol. 35, 1980.

[14] T. A. C. M. Classen and W. F. G. Mechlenbrauker, "The wigner distribution - a tool for time-frequency signal analysis: Part II: Discrete-time signals," *Phillips J. Res.*, vol. 35, 1980.

[15] A. W. Rihaczek, *Principles of High Resolution Radar*. New York: McGraw-Hill, 1969.

[16] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. ASSP-63, Dec. 1975.

[17] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-27, Feb. 1979.

[18] J. L. Ziedler *et al.*, "Adaptive enhancement of multiple sinusoids in uncorrelated noise," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-26, June 1978.

[19] J. T. Richard and J. Ziedler, "Second order output statistics of the adaptive line enhancer," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-27, Feb. 1979.

[20] N. J. Bershad *et al.*, "Tracking characteristics of the LMS adaptive line enhancer response to a linear chirp signal in noise," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-28, Oct. 1980.

[21] J. R. Glover, "Adaptive noise canceling applied to sinusoidal interference," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-25, Dec. 1977.

[22] L. J. Griffith, "A continuously-adaptive filter implemented as a lattice structure," in *Proceedings IEEE ICASSP77*, (Hartford, Conn.), May 1977.

[23] L. J. Griffith, "An adaptive lattice structure for noise-cancellation applications," in *Proceedings IEEE ICASSP78*, (Tulsa, Ok), Apr. 1978.

[24] V. U. Reddy, B. Egardt, and T. Kailath, "Optimized lattice-form adaptive line enhancer for a sinusoidal signal in broadband noise," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-29, June 1981.

[25] B. Friedlander, "A recursive maximum likelihood algorithm for ARMA line enhancement," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-30, Aug. 1982.

[26] B. D. Rao and S. Y. Kung, "Adaptive notch filtering for the retrieval of sinusoids in noise," *IEEE Transactions on Acoustic, Speech and Signal Processing*, vol. ASSP-32, Aug. 1984.

[27] A. Nehorai, "Minimal parameter adaptive notch filter with constrained poles and zeros," *IEEE Transactions on Acoustic Speech and Signal Processing*, vol. ASSP-33, Aug. 1985.

[28] W. Tong, E. I. Plotkin, and M. N. S. Swamy, "Blind deconvolution of linear system driven by non-stationary source based on almost-symmetrical time varying ARMA model," in *Proceedings ISCAS93*, June 1993.

[29] W. Tong, D. Wulich, E. I. Plotkin, and M. N. S. Swamy, "Stability of two second order time-varying constrained notch filter with externally controlled coefficients," in *Proceedings of Canadian Conference of Electrical and Computer Engineering*, (Ottawa, Canada), Sept. 1990.

[30] W. Tong, E. I. Plotkin, M. N. S. Swamy, and D. Wulich, "Closed loop self synchronized signal controlled constrained notch filter with application for rejection of arbitrary frequency variation speed sinusoidal interference," in *Proceedings ISCAS91*, (Singapore), pp. 2431–2434, June 1991.

[31] W. Tong, E. I. Plotkin, and M. N. S. Swamy, "Cross-coupled constrained null filter and separation of superimposed frequency modulated sinusoids," in *Proceedings ISCAS92*, pp. 1479–1482, May 1992.

[32] W. Tong, *Almost-Symmetrical Time-Varying ARMA Model and Its Applications for Separation of Signals with Overlapping Fourier Spectra*. PhD thesis, Concordia University, 1993.

[33] E. I. Plotkin and M. N. S. Swamy, "Structural signal processing – current trends," *The American Society of Mechanical Engineers*, 1989.

[34] N. V. Thakor, "Adaptive filtering of evoked potentials," *IEEE Trans. Biomedical Engineering*, vol. BME-34, pp. 6–12, Jan. 1987.

[35] M. V. Spreckelsen and B. Bromm, "Estimation of single evoked cerebral potentials by means of parametric modeling and Kalman filtering," *IEEE Trans. Biomedical Engineering*, vol. BME-35, pp. 691–700, Sept. 1988.

[36] D. Urbach and H. Pratt, "Application of finite impulse response digital filters to auditory brain-stem evoked potentials," *Electroencephalography and clinical Neurophysiology*, vol. 64, pp. 269–273, xxx 1986.

[37] G. Shangkai and M. H. Loew, "An autoregressive model of the BAEP signal for hearing-threshold testing," *IEEE Trans. Biomedical Engineering*, vol. BME 33, pp. 560–565, June 1986.

[38] A. M. Zayezdny, E. I. Plotkin, and Y. Cherkassky, "Signal processing using their structural properties," *Telecomm. Radio Engrg*, vol. 26, pp. 61–65, 1971. English Translation.

[39] C. D. McGillem and G. R. Cooper, *Continuous and Discrete Signal and System Analysis*. New York: Holt, Rinehart and Winston, 2nd ed., 1984.

[40] L. E. Franks, *Signal Theory*. Englewood Cliffs, NJ: Prentice Hall, 1969

[41] C. E. Shannon, "Communication in the presence of noise," *Proc. of IRE*, vol. 37, 1949.

[42] J. M. Whitaker, *Interpolatory Function Theory*. Cambridge, New York: Cambridge Trac 33 in Mathematicas and Mathematical Physics, 1935.

[43] R. A. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge University Press, 1985.

[44] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*, vol. Part I. New York: John Wiley & Sons, 1968.

[45] B. de Prony (Gaspard Riche), "Essai expérimental et analytique: sur les lois de la dilatabilité des fluides élastiques et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures," *J. Ec. Polytech*, vol. vol. 1, cahier 2, pp. 24–76, 1795.

[46] J. V. Candy, *Signal Processing: The Modern Approach*. New York: McGraw-Hill, 1988.

[47] W. Hurewicz, *Lectures on Ordinary Differential Equations*. New York: John Wiley & Sons, 1958.

[48] T. Muir, *Theory of Determinants*, vol. 1-4. New York: Dover, 1906.

[49] P. M. DeRusso, R. J. Roy, and C. M. Close, *State Variables for Engineers*. New York: John Wiley & Sons, 1965.

[50] E. I. Plotkin, A. M. Zayezdny, L. Roytman, and M. N. S. Swamy, "Simulation of modulated signals," *Journal of Flanklin Institute*, vol. 318, pp. 15–28, July 1984.

[51] B. Kuo, *Automatic Control Systems*. Englewood Cliffs, NJ: Prentice-Hall, 6th ed., 1991.

[52] D. Wulich, E. I. Plotkin, M. N. S. Swamy, and W. Tong, "PLL synchronized time-varying constrained notch filter for retrieval of a weak multiple sine signal jammed by FM interference," *IEEE Transaction on Signal Processing*, vol. ASSP-33, Nov. 1992.

[53] R. Agarwal, E. I. Plotkin, and M. N. S. Swamy, "Improved design of modified notch filters," *Int. J. of Circuit Theory and Applications*, pp. 321–334, July 1994.

[54] R. Agarwal, E. I. Plotkin, and M. N. S. Swamy, "LMS-optimal notch filters with improved transient performance," in *Proceedings MWCAS94*, (Lafayette,Louisiana), August 1994.

[55] S. Sunder and V. Ramachandran, "Efficient design of recursive filters satisfying prescribed magnitude and phase specification," in *Proceedings of IEEE Pacific Rim Conference on Communication, Computers and Signal Processing*, (Victoria, Canada), pp. 398–401, May 1993.

[56] A. Deczky, "Synthesis of recursive filters using minimum p criterion," *IEEE Trans on Audio and Electroacoustics*, vol. AU-20, pp. 257-263, Oct. 1972.

[57] P. Z. Peebles, *Probability Random Variables and Random Signal Principles*. New York: McGraw-Hill, 1987.

[58] G. L. Turin, "An introduction to matched filters," *IRE Trans Inform. Theory*, vol. IT-6, 1960

[59] L. A. Zadeh and J. R. Ragazzini, "Optimum filters for detection of signal in noise," *Proc. of IRE*, vol. 40, 1952.

[60] T. Kailath, "Correlation detection of signals perturbed by random channel," *IRE Trans. Inform. Theory*, vol. IT-, June 1960.

[61] J. G. Proakis, *Digital Communications*. New York: McGraw Hill, 1983.

[62] D. Middleton, *An introduction to Statistical Communication Theory*. New York. McGraw-Hill, 1960.

[63] M. Frish and H. Messer, "Transient signal detection using prior inforamtion in the likelihood ratio test," *IEEE Trans. Signal Processing*, vol. 41, June 1993

[64] M. J. Hinich, "Detecting a transient signal by bispectral analysis," *IEEE Trans. Acoust., Speech and Signal*, vol. ASSP-38, pp. 1277-1283, July 1990.

[65] Y. Yoganandam, R. Agarwal, E. I. Plotkin, and M. N. S. Swamy, "Automatic detection of evoked ABRs," in *ICRAMBE*, (Hyderabad, India), Jan. 1993.

[66] G. W. Wornell and A. V. Oppenheim, "Estimation of fractal signals from noise measurements using wavelets," *IEEE Trans. Signal Processing*, vol. 42, July 1992.

[67] A. Papoulis, *Signal Analysis*. New York: McGraw-Hill, first ed., 1977.

[68] J. Zhang and G. Walter, "A wavelet-based KL-like expansion for wide sense stationary random processes," *IEEE Trans. Signal Processing*, vol. 42, July 1994.

[69] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice Hall, 1986.

[70] T. Demiralp and A. Ademoglu, "Modeling of evoked potentials as decaying sinusoidal oscillations by prony method," in *Proceedings EMBC'1992*, (Paris, France), November 1992.

[71] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans of ASME Journal of Basic Engineering*, Mar. 1960.

[72] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ. Prentice-Hall, 1979.

[73] R. G. Brown, *Random Signal Analysis and Kalman Filtering*. New York: John Wiley & Sons, 1983.

[74] D. Kahaner, *Numerical Methods and Software*. Englewood Cliffs, NJ: Prentice-Hall, 1989

[75] K. J. Astrom and B. Wittenmark, *Computer Controlled Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1990.

[76] R. Agarwal, "Parameter retrieval of auditory evoked brainstem response using constrained adaptive notch filters and cross-correlation techniques," Master's thesis, Concordia University, 1991.

[77] B. A. Weber and G. L. Fletcher, "A computerized scoring procedure for auditory brainstem response audiometry," *Ear Hearing*, vol. 1, 1980.

[78] K. H. Chiappa, K. J. Gladstone, and R. R. Young, "Brainstem auditory evoked responses," *Arch Neurol.*, vol. 36, Feb. 1979.

[79] W. Woodsworth, S. Reisman, and B. Fontaine, "The detection of auditory evoked responses using a matched filter," *IEEE Trans. Biomedical Engineering*, vol. BME-30, July 1983.

[80] T. W. Picton *et al.*, "Evoked potentials: How now?," *Journal of Electrophysiological Technology*, vol. xxx, Nov. 1984.

[81] L. J. Hood and C. I. Berlin, *Auditory Evoked Potentials*. Austin, Texas: Pro-Ed, 1986.

[82] S. A. Arnold, "Objective versus visual detection of the auditory brainstem response," *Ear Hearing*, vol. 6, 1985.

[83] R. J. Salvi *et al.*, "Evoked potentials: Computer automated threshold tracking procedure using an objective detection criterion," *Ear and Hearing*, vol. 8, 1987.

[84] P. K. H. Wong and R. G. Bickford, "Brainstem auditory evoked potentials: the use of noise estimates," *Electroencephlogr. Clin Neurophysiol*, 1980.

[85] G. E. Salamon, "Electric response audiometry (ERA) based on rank correlation," *Audiology*, vol. 13, 1974.

[86] J. D. Wicke *et al.*, "On line statistical analysis of averaged evoked potentials: application to evoked response audiometry (ERA)," *Electroencephlogr Clin Neurophysiol*, vol. 44, 1978.

[87] J. L. DePalma, *Modeling of the evoked potential as a linear combination of exponentials*. PhD thesis, Colorado State University, Fort Collins, Colorado, Fall 1990.

[88] R. Agarwal, E. Plotkin, and M. N. S. Swamy, "Cascade implementation of CANFs in spectral analysis of short data auditory evoked brainstem responses," in *Proceedings ISCAS91*, (Singapore), June 1991.

[89] R. Agarwal, E. I. Plotkin, and M. N. S. Swamy, "Statistically optimal filtering of averaged ABRs," in *Proceedings of EMBC '95*, (Montreal, Canada), Sept 1995