

**OPTIMAL FINITE IMPULSE RESPONSE FILTER DESIGN
BASED ON MINIMIZING THE ITAKURA-SAITO
DISTORTION MEASURE WITH APPLICA-
TIONS TO DIGITAL SPEECH
COMMUNICATIONS**

By

CHINDAKORN TUCHINDA

**Bachelor of Engineering in Electrical Engineering
Chulalongkorn University
Bangkok, Thailand
1986**

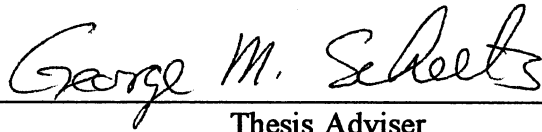
**Master of Science
Oklahoma State University
Stillwater, Oklahoma
1989**

**Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 1992**

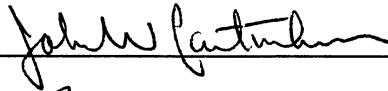
Thesis
1992D
T8880

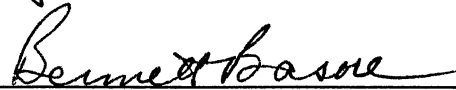
OPTIMAL FINITE IMPULSE RESPONSE FILTER DESIGN
BASED ON MINIMIZING THE ITAKURA-SAITO
DISTORTION MEASURE WITH APPLICA-
TIONS TO DIGITAL SPEECH
COMMUNICATIONS

Thesis Approved:



Thesis Adviser









Dean of the Graduate College

ACKNOWLEDGEMENTS

I wish to express sincere appreciation to my major adviser, Dr. George M. Scheets, who, despite his many responsibilities, always found the time to give me all guidance and assistance I needed. I am totally indebted to his academic and financial support throughout my graduate program. His patience, encouragement, and understanding for the last four years made working with him a pleasure.

In addition, I want to extend my thanks to my other thesis committee members, Dr. Bennette Basore, Dr. John Cartinhour, Dr. William Warde, and Dr. Jerry Johnson for their support. I am grateful to Dr. James Baker for offering me an assistantship throughout my graduate program at Oklahoma State University.

Many thanks to Miss Chwee-Ping Missy Eng for her love and support during all these years in the United State of America. I want to dedicate this thesis to my belated uncles, Mr. Ahram Kotikula, who will always be my inspirer for all my life, and Mr. Thanat Gojaseni, who encouraged me to pursue a career in Electrical Engineering.

Most of all, I would like to thank my parents, Dr. Uthai and Dr. Saisudchai Tuchinda, and my only brother, Dr. Visit Tuchinda, for their love and constant encouragement. Without such wonderful people it would not have been possible or worthwhile.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
1.1 Speech Production Mechanism and Human Auditory System	3
1.2 Frequency Domain Analysis of the Speech Signal	7
1.3 Overview	10
1.4 Summary	13
II. SPEECH DISTORTION MEASURES	14
2.1 Introduction	14
2.2 Speech Distortion Measure	16
2.2.1 Time Domain Distortion Measure	17
2.2.1.1 Signal-to-Noise Ratio	17
2.2.1.2 Mean Square Error	18
2.2.2 Spectral Distortion Measure	19
2.2.2.1 L_p Norm of the Difference of the Log Spectra	20
2.2.2.2 Itakura-Saito (IS) Distortion Measure	21
2.2.2.3 Itakura Distortion Measure	22
2.2.3 Cepstral Distortion Measure	23
2.2.4 Mean Opinion Score	24
2.2.5 Other Distortion Measure	24
2.3 Derivation of the IS Distortion Measure	25
2.4 Relationship between the IS Distortion Measure and the Information Discrimination Function	32
2.5 Literature Review	37
2.6 Contribution of this Research	42
2.7 Summary	44

Chapter	Page
III. OPTIMAL POST FILTER DESIGN	45
3.1 Introduction	45
3.2 Limitations of Realizing the Wiener by Solving the Wiener-Hopf System of Equations	47
3.3 Derivation of the Optimal IS Filter	52
3.4 Computer Simulation Results and Discussions	62
3.5 Summary	75
IV. JOINTLY OPTIMAL PRE- AND POST FILTER DESIGN	76
4.1 Introduction	76
4.2 Derivation of Jointly Optimal Pre- and Post- IS Filters	80
4.3 Computer Simulation Results and Discussions	89
4.4 Summary	98
V. REAL SPEECH SIMULATIONS	99
5.1 Computer Simulation Results and Discussions	100
5.2 Summary	139
VI. APPLICATION OF THE OPTIMAL IS FILTER IN THE DCT DOMAIN .	140
6.1 Introduction	140
6.2 Optimal IS filter in the DCT domain	143
6.3 Computer Simulation Results and Discussions	147
6.4 Summary	158
VII. CONCLUSIONS	160
7.1 Summary	160
7.2 Considerations for Future Research	164
REFERENCES	166

LIST OF TABLES

Table	Page
3.1 Comparison of the Wiener Filter Coefficients and the Optimal IS Filter Coefficients	66
4.1 Comparison of IS Distortion Measure of the Single Optimal IS Filter and Jointly Suboptimal System	97

LIST OF FIGURES

Figure	Page
1.1 Digital Speech Communication System	2
1.2 Sagittal Plane X-Ray of the Human Vocal Apparatus	5
1.3 Speech Model	5
1.4 Schematic Drawing of the Peripheral Auditory System	6
1.5 Plot of Second Formant Frequency Versus First Formant Frequency for Vowels	9
1.6 Plot of Second Formant Frequency Versus First Formant Frequency for the Diphthongs	9
2.1 Distortion Measuring System	15
2.2 Vector Quantization System	39
2.3 Filtering Scheme	43
3.1 Optimal Filtering System	46
3.2 Plot of the Forward IS Distortion Measure Versus Number of Iterations (Vary the Filter Order)	64
3.3 Plot of the Backward IS Distortion Measure Versus Number of Iterations (Vary the Filter Order)	65
3.4 Plot of the Output SNR Versus Number of Iterations (Vary the Filter Order)	68
3.5 Plot of the Forward IS Distortion Measure Versus Number of Iterations (Vary the Corrupting Noise Variance)	69

Figure	Page
3.6 Plot of the Backward IS Distortion Measure Versus Number of Iterations (Vary the Corrupting Noise Variance)	70
3.7 Plot of the Output SNR Versus Number of Iterations (Vary the Corrupting Noise Variance)	71
3.8 Comparison of the Power Spectrums of the Optimal IS Filter Output and the Wiener filter Output ($\sigma_u^2=1$)	73
3.9 Comparison of the Power Spectrums of the Optimal IS Filter Output and the Wiener filter Output ($\sigma_u^2=5$)	74
4.1 Jointly Optimal Pre- and Post-Filtering Communication System	77
4.2 Plot of the IS Distortion Measure Versus Number of Iterations (Vary the Filter Orders)	90
4.3 Plot of the Output SNR Versus Number of Iterations (Vary the Filter Orders)	91
4.4 Plot of the IS Distortion Measure Versus Number of Iterations (Vary Corrupting Noise Variances)	93
4.5 Plot of the Output SNR Versus Number of Iterations (Vary Corrupting Noise Variances)	94
4.6 Plot of the IS Distortion Measure Versus Number of Iterations for Three Combinations of Corrupting Noise Variances	95
4.7 Plot of the Output SNR Versus Number of Iterations for Three Different Combinations of Corrupting Noise Variances	96
5.1 Plot of the First 7680 Samples of Sentence One.	102
5.2 Plot of the First 7680 Samples of the Wiener Filter Output	103
5.3 Plot of the First 7680 Samples of the Optimal IS Filter Output	104
5.4 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 1$)	106

Figure	Page
5.5 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 2$)	107
5.6 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 3$)	108
5.7 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 1$)	109
5.8 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 2$)	110
5.9 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 3$)	111
5.10 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different Noise Variances	112
5.11 Plot of the First 7680 Samples of the Optimal IS Filter Output (Order = 2)	114
5.12 Plot of the First 7680 Samples of the Optimal IS Filter Output (Order = 10)	115
5.13 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different FIR Filter Orders	116
5.14 Plot of the First 7680 Samples of Sentence Two	118
5.15 Plot of the First 7680 Samples of the Wiener Filter Output	119
5.16 Plot of the First 7680 Samples of the Optimal IS filter Output	120
5.17 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 1$)	121
5.18 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 2$)	122
5.19 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 3$)	123

Figure	Page
5.20 Comparison of the Wiener filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2=1$)	124
5.21 Comparison of the Wiener filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2=2$)	125
5.22 Comparison of the Wiener filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2=3$)	126
5.23 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different Noise Variances	127
5.24 Plot of the First 7680 Samples of the Output of the Optimal IS Filter (Order = 2)	129
5.25 Plot of the First 7680 Samples of the Output of the Optimal IS Filter (Order = 10)	130
5.26 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different Filter Order	131
5.27 Plot of the First 7680 Samples of the Output of the Jointly Optimal System (Sentence One)	135
5.28 Comparison of the Jointly Optimal System and the Single Optimal System in Terms of the Autocorrelation Function Error (Sentence One)	136
5.29 Plot of the First 7680 Samples of the Output of the Jointly Optimal System (Sentence Two)	137
5.30 Comparison of the Jointly Optimal System and the Single Optimal System in Terms of the Autocorrelation Function Error (Sentence Two)	138
6.1 DCT Communication System	142
6.2 Plot of the First 7680 Samples of the IDCT of the Wiener Filter Output	148
6.3 Plot of the First 7680 Samples of the IDCT of the Optimal IS Filter Output	149

Figure	Page
6.4 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 1$)	150
6.5 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 2$)	151
6.6 Plot of the First 7680 Samples of the Optimal IS Filter ($K = 32$)	153
6.7 Comparison of the IS Distortion Measure of the Optimal IS Filter for $K = 0$ and $K = 32$	154
6.8 Comparison of the IS Distortion Measure of the Optimal IS Filter in the DCT Domain and Optimal IS Filter in the Time Domain ($\sigma^2 = 1$)	156
6.9 Comparison of the IS Distortion Measure of the Optimal IS Filter in the DCT Domain and Optimal IS Filter in the Time Domain ($\sigma^2 = 2$)	157

NOMENCLATURE

A_i	Vector of FIR filter coefficients
a_i	FIR post-filter coefficient
b_i	FIR pre-filter coefficient
$d(x,y)$	Speech distortion measure
d_{IS}	IS distortion measure
$F(\omega)$	Frequency response of the pre-filter
$G(\omega)$	Frequency response of the post-filter
$I_N(X,Y)$	Information discrimination function
J_i	Jacobian matrix at the i^{th} iteration
$L(X/\Theta)$	Log likelihood function of X given Θ
$P(\omega)$	Power spectrum of discrete-time signal
R	Autocorrelation matrix
$R(i)$	Autocorrelation of discrete-time signal
σ^2	Variance of white Gaussian noise
Θ	AR spectral model parameter
$v_x(i)$	the i^{th} DCT coefficient of $x(n)$'s
X	Vector of $x(n)$'s

CHAPTER I

INTRODUCTION

Information exchange by speech plays a crucial role in our lives. Speech conveys several kinds of acoustic information such as meaning, information as to whom is speaking, emotion of the speaker, etc.. The first item, meaning, is the most important one. It is known that acoustic transmission and reception of speech is limited for a short distance. This is because the radiated acoustic energy dies out rapidly as the distance increases. In addition, the transmission medium allows only limited variations in pressure without distorting the signal [Fla72]. Thus, the acoustic wave is not a good means for long distance speech communication.

Long distance speech communication has become a major factor in our lives since the invention of the telephone. Many researchers have tried to devise speech signal processing techniques that can yield high perceptual quality speech at the receiving end of a communication system. In the past five years, with the evolution of powerful microprocessors, the trend of using digital signal processing (DSP) techniques has become very attractive. Nowadays, a complicated on-line DSP algorithm can be computed within a few seconds or less by a high speed microprocessor.

A general digital speech communication system is depicted in Figure 1.1. A

speech signal is first sampled and quantized to obtain a digital representation of the speech sound. The quantized signal is fed into a DSP algorithm in order to either enhance the speech perceptual quality as in high quality speech signal processing, or remove the redundancy in the speech signal for compression purposes. The output of the DSP algorithm is transmitted through a communication channel by a transmitter. At the receiving end, the received signal distorted by existing noise is processed by another DSP algorithm to extract the important features of the speech signal. The output of the DSP algorithm is fed into a synthesizer to reproduce the speech sound.

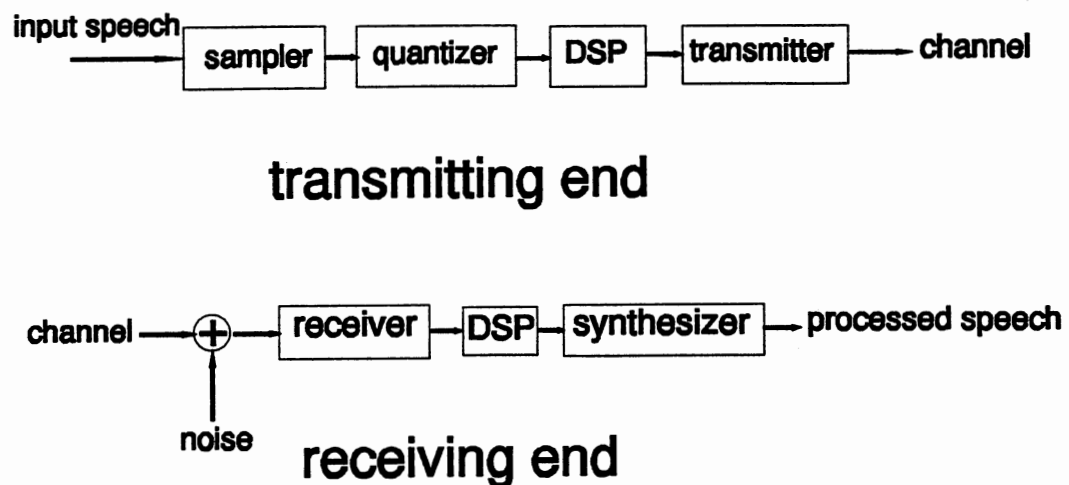


Figure 1.1. Digital Speech Communication System.

In order to be able to properly design a digital speech signal processing algorithm, it is useful to have basic understanding of both the speech production mechanism and the human auditory system which we will discuss in the following section. In section 1.2, we will see that analyzing the speech signal in the frequency domain gives us more insight as to how a human produces and perceives a speech sound. In section 1.3, we present an overview of this research. Finally, the conclusion of this Chapter is given in section 1.4

1.1 Speech Production Mechanism and Human

Auditory System

Sound is a pressure wave which consists of the vibration of molecules of an elastic medium. Sound can be considered as a physical disturbance of air particles caused by vibrating objects, e.g., vocal cords, bells, etc.. The motion of an air molecule is transmitted to adjacent air molecules where the vibration is repeated causing the sound wave to propagate. Sound waves have some properties which make them different from vibrations of other kinds as follows [Ger74].

1. Sound waves in free air propagate in three dimensions.
2. A sound wave in free air is a longitudinal wave, i.e., the motion of the air molecules is in the direction of propagation.

Human speech, which is a special kind of sound wave, can be classified into three different groups; voiced sound, fricative or unvoiced sound, and plosive sound. Each group is different in terms of which human organs are used to produce a speech

sound. The major organs involved in producing speech are illustrated in Figure 1.2. A voiced sound is produced by forcing quasi-periodic pulses of air from the lungs through the glottis, causing the vocal cords to vibrate. Unvoiced sound is generated by forming a constriction at some point in the vocal tract and forcing air through the constriction at a sufficiently high enough velocity to produce turbulence. Plosive sound results from building up pressure behind a closure and abruptly releasing it [Fla72, Rab78]. As a result, it is possible to model a speech sound by a time varying model depicted in Figure 1.3. The model consists of a sound source and time varying vocal tract filter. The output of the sound source can be considered as a fast varying portion of the speech sound, while the time varying vocal tract filter can be considered as a slow varying portion of the speech signal. Even though the speech signal is time varying, i.e., nonstationary, for a short frame of speeches, we can assume that the speech signal is stationary [Rab78].

The speech wave produced by the vocal organs propagates through the air to the ears of the listeners. At the ear, the speech wave activates the hearing organs to produce nerve impulses which are transmitted to the listeners' brain through the auditory nerve system. The human hearing organ is depicted in Figure 1.4. A sound wave causes vibration of the eardrum. This vibration is picked up by the ossicular bone of the middle ear and retransmitted to the cochlea causing motion of the basilar membrane (the cochlea is a slender, fluid filled tube divided into two chambers by the basilar membrane). This motion is sensed by the auditory nerve to recognize which sound is heard. Different speech sounds will create different patterns of the basilar

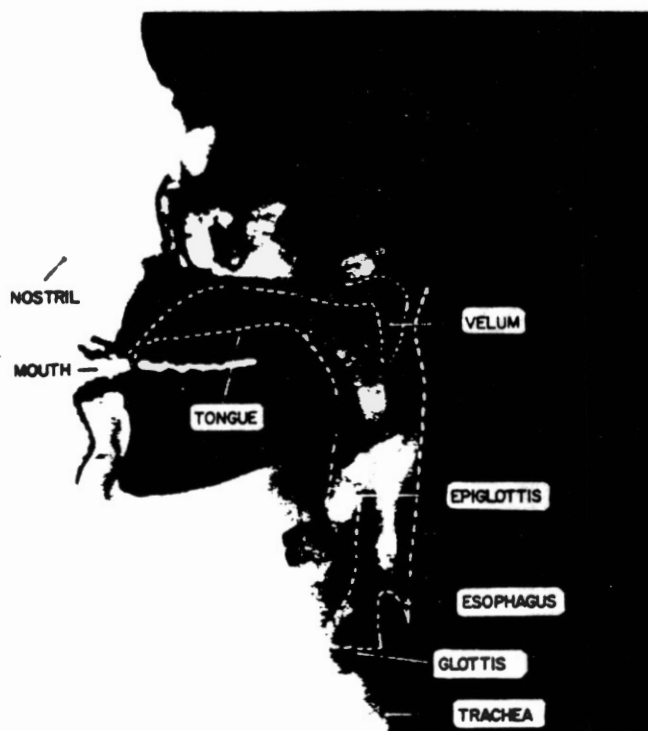


Figure 1.2. Sagittal Plane X-Ray of the Human Vocal Apparatus [Fla70].

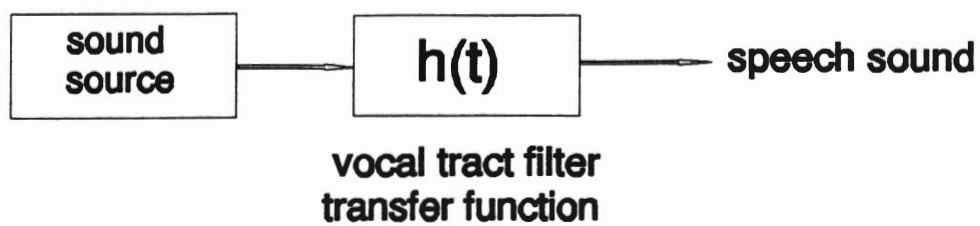


Figure 1.3. Speech Model.

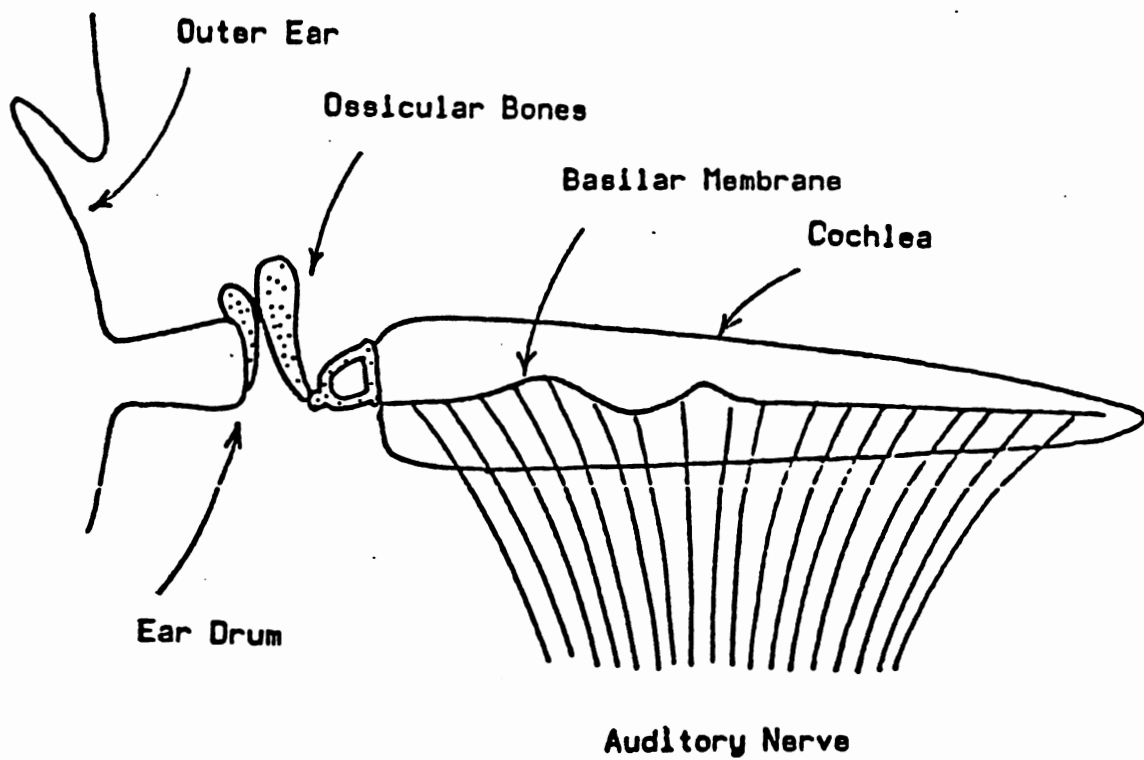


Figure 1.4. Schematic Drawing of the Peripheral Auditory System [Cal72].

membrane's motion. Bekesy [Bek60] showed that displacement of the basilar membrane is greatest at the point of resonance which corresponds to the component of the speech signal containing the highest energy level. As a result, the human auditory system is more sensitive to high energy components of speech than the lower energy ones. However, to accomplish the hearing process, both the frequencies and amplitude (or intensity) of the components of the speech must be within the limits of the response of the ear called "the hearing range". We note that limiting the range of speech frequency is equivalent to performing band-pass filtering on the received speech sound.

1.2 Frequency Domain Analysis of the Speech Signal

The behavior of speech signals can be easier to explain in frequency domain than in time domain. We saw in the previous section that the displacement of the basilar membrane is sensitive to the resonance frequency of the speech signal. Two distinct speech sounds have distinct resonance frequencies creating two distinct motions of the basilar membrane; allowing us to recognize the difference.

Most languages, including English, can be described in terms of set of distinctive sounds, or phonemes. In American English, there are 42 phonemes including vowels, diphthongs, semivowels, and consonants which depend on how the sound is generated. For example, vowels are produced by exciting the vocal tract with quasi-periodic pulses of air created by vibration of the vocal cords. Unfortunately, even the same phoneme pronounced by different speakers can yield different

sequences of what are called formant frequencies. Formant frequencies are the dominant frequencies corresponding to the resonant frequencies of the vocal tract components which characterize the phonemes. However, if the variation of the formant frequencies is not too significant, the human auditory system still can recognize which phoneme is heard. Experimental results show that only the first two or three formants are important in terms of recognizing what sound is heard while the higher formants correspond to high quality speech sound [Fla72]. Thus, two different speech sounds with close formants will sound alike.

In Figure 1.5, we show the plot of second formant frequency versus first formant frequency for vowels by a wide range of speakers. In Figure 1.6, we show the plot of the time variations of the first two formants for diphthongs. As we can see, each phoneme is categorized into different groups. The phonemes within the same group will sound alike and be recognized as the same sound. In a speech communication system, as long as the processed speech belongs to the same phoneme group as the original speech, the listener should be able to recognize which sound is transmitted. Thus, a good receiver for speech communication system should have ability to assign the received distorted speech signal to its associated phoneme group. This technique is called the nearest neighborhood system. The nearest neighborhood system is quite closely related to the maximum likelihood detection in a Gaussian process [Joh92]. The received speech signal will be assigned to the phoneme group that yields the minimum distortion measure between the original speech and the processed speech. This technique is also known as discriminant analysis [Gol78]. In

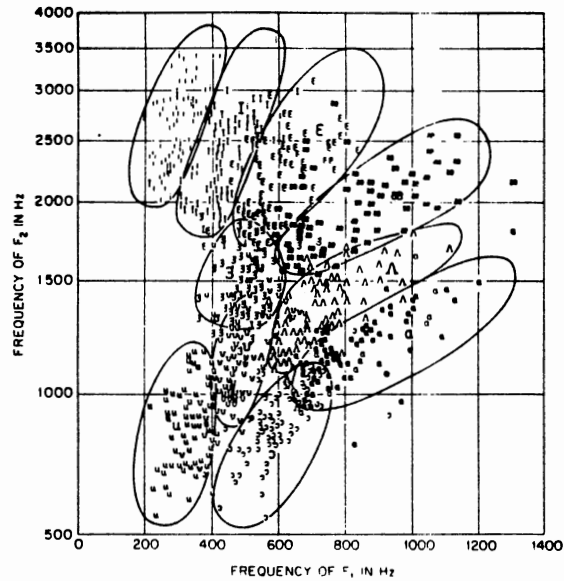


Figure 1.5. Plot of Second Formant Frequency Versus First Formant Frequency for Vowels [Pet52].

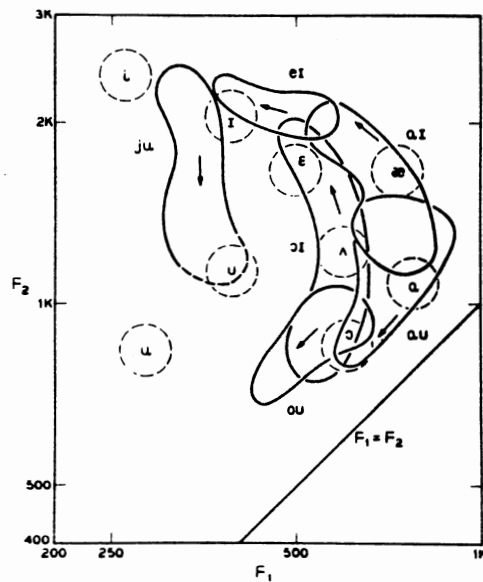


Figure 1.6. Plot of Second Formant Frequency Versus the First Formant Frequency for Diphthongs [Hol62].

discriminant analysis, several speech distortion measures can be used, for example, the mean square error (MSE), signal-to-noise ratio (SNR), the Itakura-Saito (IS) distortion measure, etc.. The most commonly used is probably the MSE due to its simplicity to implement. However, in speech recognition, the MSE does not serve well as a judgement of speech quality, since a large value of MSE does not always imply poor speech quality. In this research, we will concentrate on one special type of the speech distortion measure called the IS distortion measure. This IS distortion measure will be used as a performance index for designing an optimal finite impulse response (FIR) filter called an optimal IS filter. The IS distortion measure is closely related to the information discrimination function of two Gaussian random processes (the minimum discrimination function is a tool to measure similarity between two Gaussian random processes). By viewing the speech signal as a stationary random process (this is valid for only a short frame of speech data), we will see later that minimizing the IS distortion measure between the original speech and the processed speech is equivalent to finding the nearest neighborhood stationary random process (the processed speech) to the original speech. As a result, compared to the MSE, one would expect the IS distortion measure to perform better.

1.3 Overview

The introduction of speech distortion measure is given in Chapter 2. Several types of speech distortion measure are discussed. Later, we concentrate on the IS distortion measure. The IS distortion measure was first introduced as an error

matching function between the autoregressive (AR) spectral model and the short time spectrum of speech [Ita68, Ita70]. In addition, Itakura also showed that the maximum likelihood estimator of the AR spectral model parameters could be found by the famous linear prediction coder (LPC) analysis. We then discuss the relationship of the IS distortion measure and the information discrimination function. Pinsker [Pin64] showed that the information discrimination function is asymptotically equal to half of the IS distortion measure. Literature reviews on applications of the IS distortion measure will also be given. We then finally propose a strategy for designing a speech signal processing algorithm for perceptual purposes as follows.

For a good perceptual speech signal processing algorithm, the processed speech should satisfy the following properties.

1. The mean of the processed speech signal must be equal to the mean of the original speech signal.
2. The autocorrelation function of both signals must be equal for as many lags as possible.

The property (1) and (2) are referred as the mean and autocorrelation matching property, respectively. Preserving the autocorrelation function is equivalent to preserving the power spectrum; thus, preserving the formant frequencies.

In Chapter 3, we start with the Wiener filter and show that it does not satisfy the autocorrelation property-which is not perceptually desirable. We then derive a new optimal FIR filter obtained by minimizing the IS distortion measure between the processed speech signal and the original speech signal. It will be shown that this IS

optimal filter does satisfy the autocorrelation matching property which makes it perform perceptually better. Computer simulations are also performed to show the superiority of the optimal IS filter over the Wiener filter in terms of power spectrum matching and output SNR.

In Chapter 4, we improve the performance of the optimal IS filter by introducing another optimal FIR filter at the transmitting end. Using a pre-filter will transform the transmitting signal to be more robust to the corrupting noise existing in the communication channel. Both jointly optimal pre- and post-filters are derived. It is also shown that this jointly optimal system still satisfies the autocorrelation function matching property.

In Chapter 5, we perform real speech simulations on the optimal IS filter. Simulation results reveal that the optimal IS filter outperforms the Wiener filter not only in terms of minimizing the IS distortion measure but also autocorrelation function matching. Listening tests also show the loudness level of the optimal IS filter output is higher than that of the Wiener filter output which makes it easier to be recognized. Furthermore, we also discuss the warbling effect caused by phase distortion in the optimal IS filter.

In Chapter 6, we show the application of the optimal IS filter in the Discrete Cosine Transform (DCT) domain. It will be shown that under the DCT environment, the optimal IS filter still outperforms the Wiener filter in terms of both minimizing the IS distortion measure and autocorrelation matching. Furthermore, the optimal IS filter performs better in the DCT domain than in the time domain in terms of minimizing

the IS distortion measure. Listening tests also reveal that the warbling effect in the inverse DCT (IDCT) of the optimal IS filter output operating in DCT environment is much less than that of the optimal IS filter operating in time domain.

Finally, in Chapter 7, the results of this research to date are briefly summarized. In addition, we also discuss the future research consideration.

1.4 Summary

In this Chapter, we have discussed basic principles of the human speech production mechanism and the human auditory system. We noted that speech analysis can be more easily described in the frequency domain than time domain. In speech recognition, the first two or three formant frequencies play a dominant role in distinguishing one speech sound from another. We also discussed how speech signals can be grouped based on their corresponding sound. We noted that speech signals which belong to the same phoneme group will sound alike. We then proposed a strategy for speech signal processing in speech recognition purpose. The processed speech should satisfy the mean and autocorrelation function matching property, which is equivalent to preserving the portions of the signal which contain high energy level.

CHAPTER II

SPEECH DISTORTION MEASURES

2.1 Introduction

Consider a general communication process depicted in Figure 2.1. The unprocessed input signal, $x(n)$, is transmitted through a communication system yielding the output signal, $y(n)$. The output signal, $y(n)$, is different from the input signal, $x(n)$, due to the distortion caused by coding or the transmission process. One common question is how different the output signal is, compared to the input signal. In other words, we want to know how much the communication system distorts the input signal. The distortion measure is an objective quantity to measure the similarity between the input and output of the communication system. By objective quantity, we mean that this quantity is computed based only on the distortion caused by the coding and transmission process. However, in speech communication, the speech quality assessment involves subjective and psychological attributes of human perception, an area which mathematics cannot clearly explain. As a result, several attempts had been made to create an objective distortion measure which has an ability to predict the subjective quality as effectively as possible. Nonetheless, we should note that none of the objective distortion measures found to date can be justified as the global speech

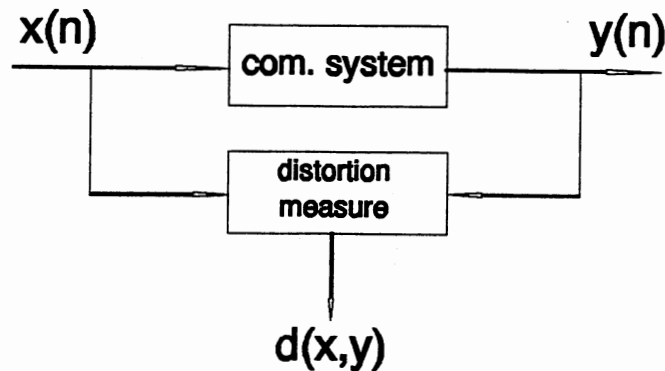


Figure 2.1. Distortion Measuring System.

quality measure [Jua84].

The purpose of this Chapter is to serve as a survey of the objective speech distortion measures. In the following section, we will review several types of distortion measures and briefly compare their advantages and disadvantages. In section 2.3, we will concentrate on one special type of spectral distortion measure called the Itakura-Saito (IS) distortion measure. This distortion measure was first introduced by Itakura [Ita68] as an "error matching function" between the speech signal and an autoregressive (AR) spectral model (also known as a discrete all-pole model). Itakura showed that if the aliasing effect is negligible the maximum likelihood estimation of the coefficients of an AR spectral model to represent the speech signal can be obtained by using the famous linear prediction coder (LPC) analysis. In other words, LPC analysis is equivalent to minimizing the IS distortion measure between the speech signal and the AR spectral model. In section 2.4, we will

discuss the relationship between the speech signal and the normalized information discrimination function. The normalized information discrimination function is a tool that measures the similarity between two Gaussian random processes. Exploring this relationship explains how the IS distortion measure relates to the nearest neighborhood system design used often in discriminant analysis. In section 2.5, we provide a literature review of how the IS distortion measure is being used. In addition, we also discuss limitations of the IS distortion measure. In section 2.6, we state the contribution of this research. Based on the knowledge gathered from the previous sections, we propose a new strategy to design a new optimal finite impulse response (FIR) filter called the optimal IS filter. This optimal filter is obtained by minimizing the IS distortion measure between the speech signal, $x(n)$, and the filter output signal, $y(n)$. Finally, the conclusion and summary is presented in section 2.7.

2.2 Speech Distortion Measures

For speech processing, the distortion measure between two frames of speech data $x(n)$ and $y(n)$, $d(x,y)$ should possess at least the following properties [Gra76]

1. $d(x,y)$ must be nonnegative, and if $x(n)=y(n)$, then $d(x,y)=0$.
2. $d(x,y)$ must be subjectively meaningful so that small and large distortion measurements correspond to good and bad subjective speech quality.
3. $d(x,y)$ must be mathematically tractable and easy to compute.

The first property assures that the valid distortion measure is positively definite. The second property links the objective measurement with human perception process.

The third property is required for practical implementation.

Note that there are some other properties that the distortion measure could satisfy, e.g., the symmetric property, the triangular inequality, etc. [Gra76]. The symmetric property is attractive since it implies that $d(x,y) = d(y,x)$. Thus, there are no restriction on either $x(n)$ or $y(n)$ to be the reference signal. Not all the distortion measures discussed in this chapter satisfy these additional properties. However, all the distortion measures satisfy the first three properties mentioned.

The objective distortion measures can be categorized into five different groups as follows.

2.2.1 Time Domain Distortion Measure

This type of distortion measure is directly evaluated from the time domain signal on a sample to sample basis. The main advantage of this type of distortion measure is tractability and ease of computation. Subjectively, the time domain distortion measure works well for the high quality (toll or near toll) speech signal. Two of the most widely used time domain distortion measures follow.

2.2.1.1 Signal-to-Noise Ratio. One of the most popular time domain distortion measures is the signal-to-noise ratio (SNR) which can be expressed as

$$\text{SNR} = \frac{\sum_{n=0}^{N-1} x^2(n)}{\sum_{n=0}^{N-1} [x(n)-y(n)]^2} \quad (2.1)$$

where N is the number of speech samples in a frame.

SNR does not fit well in terms of measuring speech subjective quality since the speech sound contains a number of pauses (silence) which degrades the SNR ability to evaluate speech quality even when the corrupting noise is small [Dim89]. In addition, since the SNR is computed on a sample by sample basis, the input signal, $x(n)$, and the output signal, $y(n)$, need to be temporally aligned or synchronized. However, some modifications can be made to increase the subjective measuring ability of the distortion measure, for instance, the segmental SNR and the frequency-weighted SNR [Jay84]. Both techniques employ the idea of introducing a weighting function on the speech samples. The erroneous portion, e.g., silence, can be suppressed by multiplication of a large quantity number. Nevertheless, these techniques are not as popular as the other objective distortion measures introduced later.

2.2.1.2 Mean Square Error. Another commonly used time domain distortion measure is the traditional mean square error (MSE) (error power or error energy). The MSE between $x(n)$ and $y(n)$ is defined as

$$\text{MSE} = E \{ [x(n)-y(n)]^2 \} . \quad (2.2)$$

where $E\{\cdot\}$ denotes the expected value.

The MSE has been successfully used in many areas of digital signal processing

including filtering, estimation, modeling, etc.. However, for a low bit rate speech system, the MSE does not appear to be subjectively meaningful [Gra80]. In particular, large distortion in the MSE does not imply poor perceptual speech quality. For example, a "shh" sound can be considered as a white process. Two completely different white processes will sound the same but may yield significantly different MSE. This effect actually stems from the property of the MSE itself. From Chapter 1, we know that human perceptual response is more sensitive to the portion of the speech which contains high energy level (this is equivalent to the formant in the frequency domain) than the portion of speech which contains lower energy level (this corresponds to the spectral valley in the frequency domain). Thus, intuitively, from the perceptual point of view, a good objective speech distortion measure would weight the high energy level portion of speech signal more than the lower energy level portion of speech signal. However, from equation (2.2), the MSE weights every speech sample equally, which is not perceptually desirable. In fact, in Chapter 3, we will show that the Wiener filter, obtained by minimizing the MSE between $x(n)$ and $y(n)$, does not preserve the second order statistical property, which makes it less subjectively meaningful.

2.2.2 Spectral Distortion Measure

Recall again from Chapter 1 that the relationship between speech signal characteristics and the human perceptual process can be more easily described in frequency domain than in the time domain. It is known that the human auditory

system is more sensitive to the spectral peaks (formants) than the spectral valleys [Fla72]. It is also useful that the first three formants are important in determining what sound is heard whereas the higher formants are necessary to produce higher quality sounds. Thus, a good spectral distortion measure should have the characteristic that spectral peaks are weighted more than spectral valleys.

We first define $P_x(\omega)$ and $P_y(\omega)$ as power spectrums of $x(n)$ and $y(n)$, respectively. Several types of the spectral distortion measures will be discussed as follow.

2.2.2.1 L_p Norm of the Difference of the Log Spectra. This is the oldest spectral distortion measure. The L_p norm of the difference of the log spectra, d_p , can be defined as [Gra76]

$$d_p = |\ln P_x(\omega) - \ln P_y(\omega)|_p, \quad (2.3)$$

where $|\cdot|_p$ denotes the L_p norm.

The typical choices of p are 1, 2, and ∞ corresponding to mean absolute, root mean square, and maximum deviation, respectively [Gra76]. We shall note that these distortion measures satisfy the symmetric and triangular property, i.e.,

$$d(x,y) = d(y,x) \quad (2.4a)$$

and

$$d(x,y) \leq d(x,z) + d(y,z) . \quad (2.4b)$$

From equation (2.3), we can see that the L_p norm spectral distortion measure

allows equal contributions from every frequency. Thus, for perceptual purposes, this distortion measure is not a very good choice. However, some improvement can be made by allowing the frequency loudness weighting function based on the transmission medium characteristic [Noc85]. However, this technique is rather complex and difficult to construct.

2.2.2.2 The Itakura-Saito Distortion Measure. The Itakura-Saito (IS) distortion measure was originally introduced as an error-matching function in the maximum likelihood estimation of an AR spectral model to represent a speech signal [Jua84, Gra80]. The forward Itakura-Saito distortion measure is defined as [Ita68, Ita70]

$$d_{IS} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P_x(\omega)}{P_y(\omega)} - \ln \frac{P_x(\omega)}{P_y(\omega)} - 1 d\omega . \quad (2.5)$$

At the present time, the IS distortion measure is the main tool to measure the similarity between two AR spectral models. As we will see in the following section, the IS distortion measure is twice the limit of the normalized discrimination function. Thus, minimizing the IS distortion measure between $x(n)$ and $y(n)$ is equivalent to finding the nearest neighborhood $y(n)$ to $x(n)$. In addition, the IS distortion measure is subjectively meaningful since it weights the spectral peaks heavier than the spectral valleys [Ita68]. For these reasons, this research is devoted to the application of the IS distortion measure in designing an optimal FIR filter. The details of the derivation and properties of the IS distortion measure will be explored extensively in the following sections.

2.2.2.3 The Itakura Distortion Measure. The Itakura distortion measure is

defined as [Ita75]

$$d_I = \ln \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P_x(\omega)/\sigma_x^2}{P_y(\omega)/\sigma_y^2} d\omega \right], \quad (2.6)$$

where σ_x^2 and σ_y^2 are average power of $x(n)$ and $y(n)$, respectively. If both $x(n)$ and $y(n)$ can modeled by two AR spectral models, i.e.,

$$P_x(\omega) = \frac{\sigma_x^2}{|A(\omega)|^2} \quad (2.7)$$

and

$$P_y(\omega) = \frac{\sigma_y^2}{|A'(\omega)|^2} \quad (2.8)$$

where

$$A(\omega) = \sum_{i=0}^p a_i e^{-j\omega i} \quad (2.9)$$

and

$$A'(\omega) = \sum_{i=0}^{p'} a'_i e^{-j\omega i} \quad (2.10)$$

and $|\cdot|^2$ denotes magnitude square. Then, it can be shown that the Itakura distortion measure is the gain-optimized version of the IS distortion measure. Furthermore, if $\sigma_x^2 = \sigma_y^2$, equation (2.6) becomes

$$d_I = \ln \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{A(\omega)}{A'(\omega)} \right|^2 d\omega \right]. \quad (2.11)$$

Equation (2.11) is also known as the log likelihood ratio distortion measure since for a Gaussian process and large sample size it does approximate a likelihood ratio of two AR spectral models [Ita75]. The Itakura distortion measure has been used successfully in speech recognition problems, especially to generate a codebook for vector quantization design since in vector quantization, the speech signal is modeled as an AR spectral model [Noc85]. However, for general cases, where $x(n)$ and $y(n)$ cannot be represented by AR spectral models, the IS distortion measure is preferable since as we will see later that the IS distortion measure closely relates to the information discrimination function used in the nearest neighborhood system while the Itakura distortion measure does not.

2.2.3 Cepstral Distortion Measure. The cepstrum, $\{c_k\}$, of a sequence $x(n)$ is the inverse Fourier transform of the logarithm power spectrum of $x(n)$ [Rab78],

$$c_k^x = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[P_x(\omega)] e^{j\omega k} d\omega. \quad (2.12)$$

The cepstral distortion measure is defined as [Gra76]

$$d_{cep} = \sum_{i=-L}^L (c_i^x - c_i^y)^2. \quad (2.13)$$

where L is any large integer number that assures d_{cep} to be positively definite.

This distortion measure was originally introduced as an estimate of d_2 distortion

measure without a DFT operation [Gra76]. Thus, the property of the cepstral distortion measure is very similar to those of d_2 .

2.2.4 Mean Opinion Score (MOS). The mean opinion score is a quantifier of subjectively rated speech quality computed by averaging the individual opinion scores from a sample of listeners. The five point opinion score, 1, 2, 3, 4, and 5, represents the quality scale, excellent, good, fair, poor, and unsatisfactory, of the speech signal. The major difficulty of this scoring system is due to the fact that listeners may occasionally rank a very slightly impaired stimulus higher than the original. In addition, the same experiment tested at different time always yields different MOS values [Jay84]. However, one may alleviate the problem by creating a distribution function of the MOS values and making use of the mean or median of such distribution function.

2.2.5 Other Distortion Measure

There are several other distortion measure introduced in the past few years. In speech recognition systems, some of them have even been reported to be more subjectively meaningful than the distortion measures discussed previously. Some of these distortion measures are weighted likelihood ratio, weight slope metric distortion measure, etc. [Noc85]. We know from Chapter 1 that the human auditory system is sensitive to the resonance frequencies (formants) in terms of recognizing what sound is heard. These distortion measures involve using a weighting function to weight the high energy level portion of a speech signal heavier than others. We also know that

two different speech sounds will have distinct set of the formants, as a result, requires two distinct weighting functions. To implement these types of distortion measure requires knowledge of what speech sound has been received in advance. However, in general cases, e.g., designing a speech optimal filter, such knowledge cannot be acquirable, i.e., we do not know which sound has been received. For these reasons, these distortion measures do not receive as much practical used as the IS distortion measure.

2.3 Derivation of the IS Distortion Measure

It is well known that an AR spectral model can be effectively used to represent a speech signal. The main problem is how to determine the best suited AR spectral model from all possible AR spectral models. The motivation of this section is to find the best AR spectral model in the maximum likelihood sense to represent the speech signal. Note that the following work has been mainly done by Itakura [Ita68, Ita70]. We start by assuming that the power spectrum of a speech signal, $P_x(\omega)$, can be modeled by a product of two components, a slowly varying (or periodicity in the case of voiced speech) positive spectrum envelope, $P_1(\omega)$, and a rapidly varying spectrum, $P_2(\omega)$, i.e.,

$$P_x(\omega) = P_1(\omega)P_2(\omega) . \quad (2.14)$$

$P_1(\omega)$ can be considered as the spectrum of the human vocal tract response and $P_2(\omega)$ corresponds to the spectrum of the source of excitation [Rab75]. The source of excitation for the voice sound is at the glottis and consists of broad-band quasi-

periodic puffs of air produced by the vibrating vocal cords. For unvoiced sound, the source of excitation can be either a turbulent quasi-random blow at the point of closure for the sound like "s" or a rapid release of the air pressure built up behind the total constriction for the sound like "p".

Itakura [Ita68, Ita70] assumes that $P_x(\omega)$ is uniform or flat spectrum. Note that this argument is true only for unvoiced speech since for voiced speech the source of excitation is periodic in nature [Rab78]. However, Itakura [Ita70] claims that a voiced sound can be expressed by a function of a periodic pulse train but with various kinds of variations such as variation of the pulse interval, the change of the pulse shape, etc.. Thus, the following analysis still can be applicable to the voiced speech.

We define a short segment of length N of a speech signal as

$$\mathbf{X} = [x(0), x(1), \dots, x(N-1)] . \quad (2.15)$$

Viewing \mathbf{X} as a random Gaussian process (note that the speech distribution function is found to be close to a Gamma or Laplacian distribution [Rab78], the goal is to model $P_x(\omega)$ by an AR spectral model of order p whose power spectrum is $P_y(\omega)$ expressed as

$$P_y(\omega) = \frac{\sigma^2/2\pi}{\left| \sum_{i=0}^p a_i e^{j\omega i} \right|^2} \quad (2.16)$$

$$= \frac{\sigma^2/2\pi}{d_0 + 2 \sum_{k=1}^p d_k \cos(k\omega)} \quad (2.17)$$

where $a_0=1$,

$$d_0 = \sum_{k=0}^p a_k^2 \quad (2.18a)$$

and

$$d_i = \sum_{k=0}^p a_k a_{k+i}, \quad 1 \leq i \leq p. \quad (2.18b)$$

Equation (2.16) is equivalent to modeling a speech frame by an AR spectral model of order p (AR(p)) driven by a white noise of variance $\sigma^2/2\pi$. Without losing any generality, we assume that $x(n)$ has zero mean. We then introduce a set of parameters

$$\Theta = [\sigma^2, a_0, a_1, \dots, a_p] . \quad (2.19)$$

Thus, the goal is to find the maximum likelihood estimator of Θ that represents a speech segment X . Under the Gaussian assumption, we can write the probability density function of X as

$$p(X) = \frac{1}{(2\pi)^{N/2} |R_x|^{1/2}} \exp \left[-\frac{1}{2} X R_x^{-1} X \right] \quad (2.20)$$

where R_x is the covariance matrix of X . When the probability density of X is given by equation (2.20), the conditional probability density function of X given Θ , $p(X/\Theta)$ can be written as [Ita70]

$$p(X/\Theta) = \frac{C}{(2\pi\sigma^2)^{N/2}} \exp \left[\frac{Q(X)}{2\sigma^2} \right], \quad (2.21)$$

where

$$C = |\sigma^2 \mathbf{R}_x^{-1}|^{1/2}, \quad (2.22)$$

and

$$Q(\mathbf{X}) = d_0 \sum_{i=0}^{N-1} x^2(i) + 2 \sum_{k=1}^p d_k \sum_{i=0}^{N-k-1} x(i)x(i+k). \quad (2.23)$$

We then introduce a short time autocorrelation function defined as

$$\hat{f}_x(j) = \frac{1}{N} \sum_{i=0}^{N-j-1} x(i)x(i+j). \quad (2.24)$$

Thus, $Q(\mathbf{X})$ can be written as

$$Q(\mathbf{X}) = N d_0 \hat{f}_x(0) + 2N \sum_{k=1}^p d_k \hat{f}_x(k). \quad (2.25)$$

From equation (2.21), we can find the log likelihood function of $p(\mathbf{X}/\Theta)$ as

$$L(\mathbf{X}/\Theta) = \ln C - \frac{N}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} Q(\mathbf{X}). \quad (2.26)$$

Itakura [Ita68, Ita70] showed that if $N \gg p$, the term $\ln C$ in the right-handed side of equation (2.26) is negligible. Thus, equation (2.26) can be approximated as

$$L(\mathbf{X}/\Theta) = \frac{N}{2} \left[\ln(2\pi\sigma^2) - \frac{1}{\sigma^2} \{d_0 \hat{f}_x(0) + 2 \sum_{k=1}^p d_k \hat{f}_x(k)\} \right]. \quad (2.27)$$

From equation (2.17), it can be shown that

$$d_i = \frac{\sigma^2}{(2\pi)^2} \int_{-\pi}^{\pi} \frac{\cos(\omega i)}{P_y(\omega)} d\omega. \quad (2.28)$$

Taking the natural logarithm of both sides of equation (2.16), we get

$$\ln P_y(\omega) = \ln \left[\frac{\sigma^2}{2\pi} \right] - \ln \left| \sum_{i=0}^p a_i e^{j\omega i} \right|^2 . \quad (2.29)$$

Assuming that the AR spectral model has all its roots inside the unit circle, it was shown that [Mar76]

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln \left| \sum_{i=0}^p a_i e^{j\omega i} \right|^2 d\omega = 0 . \quad (2.30)$$

Using equation (2.30), integrating both sides of equation (2.29) yields

$$\int_{-\pi}^{\pi} \ln P_y(\omega) d\omega = 2\pi \ln \left[\frac{\sigma^2}{2\pi} \right] . \quad (2.31)$$

Using equation (2.28) and (2.31) in equation (2.27), it is shown in [Ita70] that

$$L(\mathbf{X}/\Theta) = -\frac{N}{2} \left[\ln(2\pi)^2 + \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln P_x(\omega) + \frac{P_x(\omega)}{P_y(\omega)} d\omega \right] , \quad (2.32)$$

where $P_x(\omega)$ is the short-time power spectrum of the speech signal, $x(n)$, which can be defined as

$$P_x(\omega) = \frac{1}{2\pi} \hat{r}_x(0) + \frac{1}{\pi} \sum_{k=1}^{N-1} \hat{r}_x(k) \cos(k\omega) \quad (2.33)$$

$$= \frac{1}{2\pi N} \left| \sum_{m=0}^{N-1} x(m) e^{j\omega m} \right|^2 . \quad (2.34)$$

To obtain the maximum likelihood estimator of Θ , one needs to find the Θ

which maximizes equation (2.32) for a given X . This is equivalent to minimizing the function [Ita70]

$$F = \frac{1}{2\pi} \int_{-\pi}^{\pi} V(\omega) - \ln V(\omega) - 1 d\omega , \quad (2.35)$$

where

$$V(\omega) = \ln \frac{P_x(\omega)}{P_y(\omega)} . \quad (2.36)$$

Thus, maximizing the log likelihood function (equation (2.32)) is in fact equivalent to minimizing the IS distortion measure between a AR spectral model spectrum, $P_y(\omega)$, and the short time spectrum of the speech signal, $P_x(\omega)$. In fact, it is also shown in [Ita70] and [Mar76] that, for a band limited speech spectrum where aliasing effect is negligible, the LPC which minimizes the residue energy is equivalent to minimizing the IS distortion measure between the AR spectral model spectrum and the speech spectrum.

In LPC analysis, we want to model a speech sequence, $x(n)$ of length N by another sequence $y(n)$ in which $y(n)$ can be expressed as [Kay88]

$$y(n) = -\sum_{i=1}^p a_i x(n-i) \quad (2.37)$$

where p is the LPC order.

Such $y(n)$ can be obtained by minimizing the MSE between $x(n)$ and $y(n)$ [Mar76]. In other words, we want to minimize the residual energy, E_n , which can be defined as

$$E_n = \sum_{n=-\infty}^{\infty} e^2(n) , \quad (2.38)$$

where

$$e(n) = x(n) - y(n) \quad (2.39)$$

$$= \sum_{i=0}^p a_i x(n-i) \quad (2.40)$$

and $a_0 = 1$.

For an alternative approach to obtain the maximum likelihood estimator of Θ , Itakura and Saito (1970) start with an assumption that for $N \gg p$, the joint density function of the $x(n)$, $n = 0, \dots, N-1$, in a speech segment X can be approximated by

$$p(x(0), x(1), \dots, x(N-1)) = (2\pi\sigma^2)^{N/2} \exp[-\beta/2\sigma^2] \quad (2.41)$$

where

$$\beta = \sum_{n=-\infty}^{\infty} \left[\sum_{i=0}^p a_i x(n-i) \right]^2 . \quad (2.42)$$

We note that from equation (2.38) and (2.40), β in equation (2.42) is in fact the residue energy. Furthermore, to find a set of a_i , $i = 1, \dots, p$ which maximizes equation (2.41) is equivalent to finding the maximum likelihood estimate of a_i . Thus, the maximum likelihood estimator of Θ can be found by minimizing the residue energy in equation (2.42). We conclude that LPC analysis is equivalent to finding the spectral AR model that minimizes the IS distortion measure between $x(n)$ and $y(n)$.

So far, we have shown that the IS distortion measure is closely related to the

likelihood function of Θ given X . The maximum likelihood estimator of Θ can be obtained by minimizing the IS distortion between the speech spectrum and the AR spectral model spectrum. Furthermore, such Θ can alternatively be found by using the famous LPC analysis. However, nothing in the world is perfect. One should remember that the above analysis is based on the following assumptions.

1. The speech signal is Gaussian distributed. However, the distribution function of the human speech has been experimentally found to be closer to Gamma or Laplacian distribution function [Rab78].

2. The speech signal can be modeled by an AR spectral model driven by a white noise sequence. It is known that this model is valid for only unvoiced sound. For voiced sound, the driving sequence should be a periodic impulse train with the period corresponding to the human pitch period.

3. We assume that aliasing effects and quantization noise are negligible.

In the following section, we will look at the IS distortion measure from the information theory point of view. It will be shown that the IS distortion measure is asymptotically equal to twice the information discrimination function which is used to measure the similarity between two Gaussian random processes.

2.4 Relationship between the IS Distortion Measure and the Information Discrimination Function

Discriminant analysis and classification are multivariate techniques concerned with separating distinct sets of objects (or observations) and with allocating new

objects (or observations) to previously defined groups [Joh92]. To separate such distinct sets of objects, one tries to find a "discriminant" or information discrimination function whose numerical values are such that the collections are separated as much as possible. Given previously defined groups, a new object (or observation) will be assigned the group that yields the minimum information discrimination function. Such a technique is also known as the nearest neighborhood system. In this section, we will show that the IS distortion measure is closely related to the information discrimination function. In fact, for large samples of observation, the information discrimination function is asymptotically equal to half of the IS distortion measure. Recall from Chapter 1, the speech sounds contained in the same neighborhood will sound alike. Thus, the IS distortion measure is a very good choice to measure the quality of speech sound perceptually. As a result, this will lead us to a strategy to design a new optimal FIR filter for perceptual speech processing.

Without losing any generality, consider two zero-mean Gaussian, $x(n)$ and $y(n)$, processes with power spectrum, $P_x(\omega)$ and $P_y(\omega)$, respectively. We again consider a segment of length N of $x(n)$ and $y(n)$. Thus, we define vectors

$$\mathbf{X} = [x(0), x(1), \dots, x(N-1)] \quad (2.43a)$$

and

$$\mathbf{Y} = [y(0), y(1), \dots, y(N-1)] . \quad (2.43b)$$

Assume $x(n)$ and $y(n)$ are real. Thus, their probability density functions (PDF) of the segments \mathbf{X} and \mathbf{Y} are denoted by $p_x(\mathbf{X})$ and $p_y(\mathbf{Y})$, respectively. Since $x(n)$ and

$y(n)$ are zero-mean Gaussian processes, they can be completely described by their covariance matrices, R_X and R_Y , respectively. We shall put a constraint on Y such that R_Y is positive definite and satisfies the following property.

$$r_y(i,j) = r_x(i,j) , 0 \leq i \leq p , \quad (2.45)$$

where $r(i,j)$ is an element at row i and column j of the covariance matrix R and p is an integer less than N . We also note that $r(i,j)$ is in fact the autocorrelation function.

Thus, we can rewrite equation (2.45) as

$$R_x(i) = R_y(i) , 0 \leq i \leq p , \quad (2.46)$$

where $R_x(i)$ and $R_y(i)$ is the autocorrelation function of $x(n)$ and $y(n)$, respectively.

We note that if any positively definite R_Y satisfies equation (2.45) or (2.46), then R_Y is said to be an extension of R_X [Gra76].

The information discrimination function, $I_N(X,Y)$, of two Gaussian processes can be written as [Kul59]

$$I_N(X,Y) = \frac{1}{2} \ln \left[\frac{\det R_X}{\det R_Y} \right] + \frac{1}{2} \text{tr} \{ R_X R_Y^{-1} \} - \frac{N}{2} , \quad (2.47)$$

where $\det\{\bullet\}$ denotes the determinant of the matrix and $\text{tr}\{\bullet\}$ denotes the trace of the matrix. It is shown in Pinsker [Pin64] that this information discrimination function has a limit

$$I(X,Y) = \lim_{N \rightarrow \infty} I_N(X,Y) \quad (2.48)$$

$$= \frac{1}{2} d_{\text{IS}}(\mathbf{X}, \mathbf{Y}) . \quad (2.49)$$

Thus, the IS distortion measure is exactly twice the asymptotically information discrimination function under a Gaussian assumption. From the knowledge of the cluster analysis [Joh92], we are seeking a segment \mathbf{Y} that is closest to segment \mathbf{X} in the nearest neighborhood sense. Thus, minimizing the IS distortion measure between speech segment \mathbf{X} and \mathbf{Y} is equivalent to finding the nearest neighborhood AR spectral model of all possible AR spectral models.

In LPC analysis, given a speech segment \mathbf{X} of length N , we wish to represent a speech segment \mathbf{X} by an AR spectral model of order p which produces an estimated segment \mathbf{Y} . It is shown in [Mar76] that the LPC satisfies the autocorrelation function matching property (equation (2.46)).

Note that equation (2.46) is equivalent to equation (2.45). We also showed previously that the LPC analysis which minimizes the MSE between $x(n)$ and $y(n)$ is equivalent to finding the maximum likelihood estimator of AR spectral model parameters to represent a speech segment \mathbf{X} . In other words, we are seeking an optimal segment \mathbf{Y} to represent a segment \mathbf{X} in the sense of minimizing the MSE between $x(n)$ and $y(n)$ under constraint of equation (2.45). As a result, the LPC analysis is equivalent to minimizing the IS distortion measure between $x(n)$ and $y(n)$, or finding the nearest neighborhood segment \mathbf{Y} to represent speech segment \mathbf{X} . Thus, if aliasing is negligible, the LPC analysis is a good choice to model the speech segment.

The use of the information discrimination function can be extended to non-Gaussian data. Consider equation (2.46), it is obvious that $I_N(X,Y)$ is minimized if $R_x = R_y$. Thus, for the non-Gaussian data, to represent a speech segment X , we are seeking a segment Y which satisfies the following conditions.

$$1. E\{y(i)\} = E\{x(i)\}, i = 0,1,\dots,N-1. \quad (2.50)$$

$$2. R_y(i) = R_x(i), 0 \leq i \leq p, \quad (2.51)$$

where p is an integer greater than or equal to zero but less than or equal to N .

Equation (2.50) and (2.51) imply that to obtain the optimal segment Y one should maintain the first order random variable characteristic and preserve the second order characteristic as much as possible (up to p lags). We shall keep in mind these two equations as our strategy for designing an optimal FIR filter that has a better perceptual quality than the optimal FIR filter obtained by minimizing the MSE between $x(n)$ and $y(n)$ (called the Wiener filter). In other words, a good perceptual FIR filter should satisfy equation (2.50) and (2.51). It will be shown in the following Chapter that even though the LPC is equivalent to minimizing the IS distortion measure if the aliasing effect is negligible, this will not be the case for the Wiener filter. In other words, the Wiener filter does not satisfy the autocorrelation function matching property. Hence, filtering a speech signal by the Wiener filter is not equivalent to finding the nearest neighborhood segment Y to represent the segment X . In the next Chapter, we propose a new optimal FIR filter obtained by minimizing the IS distortion measure between $x(n)$ and $y(n)$ called an optimal IS filter. It will be

shown that this new IS filter satisfies the autocorrelation function matching property. As discussed in Chapter 1, the human auditory system is sensitive to the format frequency. Thus, the new optimal FIR filter is expected to perform perceptually better.

2.5 Literature Review

Optimal digital signal processing system has received great attention in the past twenty years. A process is optimal in the sense that the output is obtained by minimizing an objective function. The objective function may be a distortion measure described in the previous section. In speech processing, one of the most common areas happens to be the modeling of a speech signal by a linear model such as AR, or autoregressive moving average (ARMA) spectral model. In this case, the main application of the spectral distortion measure is probably in vector quantization design. In the vector quantization design, instead of transmitting the speech signal through a communication system in a sample by sample basis, we instead transmit only the AR spectral parameters Θ as discussed in the previous section [Gra84]. It is known that the typical order of the AR spectral model to represent the speech signal is about 10 to 12 [Rab75, Rab78]. Thus, a significant amount of data rate reduction can be accomplished. At the receiving end, the speech signal can be synthesized by driving the received AR spectral parameters by either a white noise sequence for the unvoiced sound or a periodic impulse for the voiced sound. However, the corrupting noise in the communication channel causes the received AR spectral parameters to deviate; thus, the synthesized sound is distorted. To alleviate this problem, research has

showed that a sufficient English sound can be reproduced within a finite number of the AR spectral parameters called a codebook. There are several methods used to generate a codebook, for example, k-mean algorithm [Lin80], split-mean algorithm [Gra80], etc.. The training speech signal is processed on a frame by frame basis. The commonly used frame size is 128 or 256 samples/frame. The AR spectral parameters for each frame are computed via LPC analysis. Once the new AR parameters are computed, the codebook is generated by grouping all the AR spectral parameters from the training speech based on how similar they are. The most commonly used similarity measures are the IS distortion measure and the Itakura distortion measure.

The basic function of a vector quantizer is depicted on Figure 2.2. A speech frame is fed into the LPC to compute the AR spectral parameters. The computed AR spectral parameters are compared with those contained inside the codebook. The AR spectral parameters which yield the nearest neighborhood distance to the computed parameters will be selected and transmitted through the communication channel. Again, the most widely used nearest neighborhood distance is the IS distortion measure. At the receiving end, the received parameters are distorted due to corrupting noise. The received parameters are then compared with all possible parameters inside the codebook. As in the transmitting end, the parameters that yield the smallest distortion measure (IS distortion measure for most case) to the received parameters will be used to resynthesize the speech sound. For further details of vector quantization design, the reader may refer to [Buz80] and [Gra84].

Note that the main purpose of the vector quantization design is to compress the

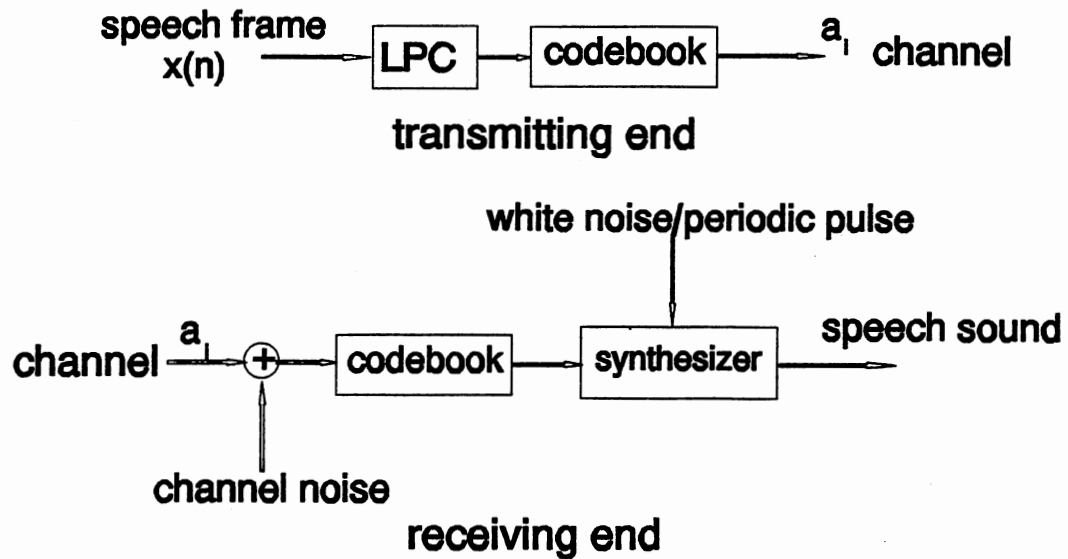


Figure 2.2. Vector Quantization System.

data rate, not to maintain high quality speech sound. The first efficient vector quantization design was introduced by [Lin80]. The design was based on the Lloyd's optimal quantizer design algorithm [Llo57]. The distortion measure used in this design is the square-error distortion measure. Compared to a simple pulse code modulation (PCM) system, the vector quantization design yields significant reductions to the data rate and fairly acceptable speech quality. Buzo [Buz80] designed a vector quantizer using the IS distortion measure as a similarity measure. For a very low data rates and a certain level of average distortion, the vector quantizer yields approximately 15 to 20 fewer bit errors per frame than the optimized scalar quantizer.

At that time, the main difficulty of the vector quantization design was the hardware implementation. Thus, the majority of the research was devoted to designing more compact codebook and more efficient search algorithms to improve the hardware speed. In terms of improving the speech quality, Chu [Chu82] proposed a technique called weighted Itakura-Saito distortion measure. This technique employed the idea that the human auditory system is more sensitive to the lower frequency band than the higher frequency band [Mak75]. The speech signal is first split into two frequency bands, 0-5 kHz and 5-10 kHz. Then, each frequency band will be modeled by the LPC where more poles will be used to represent the signal in the lower band than in the higher band. The IS distortion measure used in each band is weighted in way that it is more sensitive to the frequency band where it is used. This technique was also extended to other distortion measures, e.g., the Itakura distortion measure [Soo88, Li89]. For the moderate SNR condition, the perceptual quality (in terms of recognizing a specific word) of the Itakura distortion measure is very similar to the result obtained by the IS distortion measure. We note that all of these techniques employ very similar ideas to the technique in [Mak75]. Another approach to improve the speech quality in vector quantization design is a Fourier transform vector quantization [Cha87]. In this technique, the input speech signal is first transformed into the frequency domain via the use of discrete fourier transform (DFT). Following that, a regular vector quantizer can be employed with the addition of an inverse DFT (IDFT) at the receiving end. The distortion measure used here is simply the weighted MSE between the frequency domain of the speech signal and the frequency domain of

the estimate. The weighting value varies from frequency to frequency depending on how much the human ear corresponds perceptually. As a result, this design provides better subjective quality than ordinary vector quantization. However, the system is much more complex, as it requires two separate quantizers to take care of both real and imaginary parts of the DFT speech signal.

Recently, El-Jaroudi [Elj87, Elj88, Elj89, Elj91] proposed a new modeling technique called discrete all-pole (DAP) modeling. In DAP modeling, a speech signal is modeled by an AR spectral model which minimizes the discrete version of the IS distortion measure between the speech spectrum and the spectrum of the AR spectral model. This technique yields better spectrum fitting than the LPC technique, providing better subjective quality speech sound. Recall from the previous sections that the LPC was derived under the assumption that the sampling rate is high enough that the aliasing effect is negligible. However, for a high quality speech sound, such aliasing sometimes cannot be neglected. El-Jourdi showed that minimizing this IS distortion measure is equivalent to matching the aliased autocorrelation function of the original speech signal with the autocorrelation function of the AR spectral model aliased in the same manner up to p (order of the AR spectral model) lags. In addition, DAP modeling does have a strong relationship with the continuous spectrum linear prediction (LP) modeling [Mak75]. In fact, the DAP reduces to the LP for the continuous spectrum case but LP does not reduce to DAP for the discrete spectrum case. Thus, the LP is in fact the special case of DAP as the number of spectral points goes to infinity. We note that DAP modeling is simply an autocorrelation function

matching algorithm which is equivalent to equation (2.51). Thus, the DAP does agree with our previous conclusion that, for speech signal processing, a good perceptual DSP technique should satisfy equation (2.50) and (2.51).

2.6 Contribution of This Research

Another big area of digital signal processing is optimal digital filter design. The general digital filtering system is depicted in Figure 2.3. A receiving signal, $r(n)$, consists of a message signal, $x(n)$, corrupted by an additive white Gaussian noise, $u(n)$. The idea is to design an optimal filter which yields the output signal, $y(n)$, closest to the message signal, $x(n)$, in some sense. In other words, we are required to find an optimal filter which minimizes the distortion measure between $x(n)$ and $y(n)$, $d(x,y)$. The most commonly used distortion measure is the MSE. The optimal filter which minimizes the MSE between $x(n)$ and $y(n)$ is called the Wiener filter [Orf90]. In the past twenty years, researchers have devoted themselves to implementing the casual part of the Wiener filter. One of these techniques is the famous least mean square (LMS) algorithm [Wid84]. The LMS algorithm is successfully used to implement the echo canceler and many control algorithms. Another widely used technique is to solve the Wiener-Hopf system of equations [Orf90]. This technique has been successfully used in many areas, e.g., in digital communication system etc.. The main attraction of this technique is its ease of implementation. The autocorrelation function used in the Wiener-Hopf equation can be accurately estimated from the real time signal via many available autocorrelation function estimator [Kay88].



Figure 2.3. Filtering Scheme.

However, as mentioned in the previous section, the MSE is designed for general purposes not for the speech perception purpose. One distinct aspect is that the LPC does possess the matching autocorrelation function between the original speech signal and the estimate signal property, while the solution of the Wiener-Hopf system of equations does not as we will see in the following Chapter. Thus, one should expect that this technique is not optimal in the perceptual sense.

Up to the present, there is no optimal digital filter specifically designed for the speech perception purpose. The motivation of this research is to propose a technique for designing an optimal finite impulse response (FIR) filter for high quality speech sound. This technique is accomplished by minimizing the IS distortion measure between the original speech sound and the filter output. It will be shown later that this technique is simply equivalent to matching the autocorrelation function of the original speech with the autocorrelation function of the estimated speech up to p lags.

2.7 Summary

In this Chapter, we have discussed several speech distortion measures. We also briefly discussed some of their advantages and disadvantages. Recall in Chapter 1 that speech sounds which are in the same neighborhood will perceptually sound alike.

With this motivation, a good speech distortion measure should possess the neighborhood distortion measure minimizing property. An optimal speech processor should perform in the manner of finding the estimate which yields the minimum neighborhood distortion measure. One special type of distortion measure called the IS distortion measure was investigated in depth. We showed that this minimizing the IS distortion measure between a speech segment X and an estimate segment Y is asymptotically equivalent to finding the nearest neighborhood segment Y to represent a speech segment X . In the modeling problem, one can easily accomplish this by using the LPC analysis. However, in the filtering problem, this is not the case as we will see in the following Chapter. The Wiener filter which is closely related to the LPC analysis does not possess the autocorrelation matching property as the LPC does. Thus, the Wiener filter is not optimal in the perceptual sense. In the following chapter, we will show this limitation of the Wiener filter in detail. We also propose a new optimal FIR filter called an optimal IS filter obtained by minimizing the IS distortion measure between the speech signal, $x(n)$, and the estimated signal, $y(n)$. We will see the IS filtering technique possesses the autocorrelation function matching property while the Wiener filter does not.

CHAPTER III

OPTIMAL POST FILTER DESIGN

3.1 Introduction

The classical Wiener filter has been playing a major role in signal processing and communications since the 1950's. One of the most widely used methods is to realize the causal part of the Wiener filter with a Finite Impulse Response (FIR) filter. This method is attractive due to its ease of implementation. The coefficients of the FIR filter are obtained by solving the Wiener-Hopf system of equations which can easily be done by using Levinson algorithm [Orf90].

As discussed in the previous Chapter, the Wiener filter is designed for general, not for human perceptual, purposes. We will show in the following section that unlike the LPC, the Wiener filter does not possess the autocorrelation function matching property. Thus, it is not optimal in the perceptual sense. We will then propose a new optimal FIR filter called the optimal IS filter. This new optimal FIR filter does possess the autocorrelation function matching property while the Wiener filter does not. The optimal system is depicted in Figure 3.1. The optimal IS filter can be found by minimizing the backward IS distortion measure between the processed signal, $y(n)$, and the original signal, $x(n)$. We will show that this technique is equivalent to

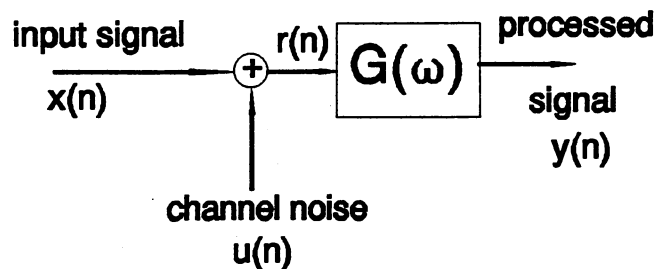


Figure 3.1. Optimal Filtering System.

matching the aliased version of the Wiener filter frequency response with the magnitude square of the optimal IS filter frequency response aliased in the same manner.

In the next section, we note the limitations of the Wiener filter by showing that the Wiener filter does not possess the autocorrelation function matching property. In section 3.3, we derive a set of normal equations for an optimal IS filter. From the set of normal equations, we will show that the optimal IS filter does possess the autocorrelation function matching property which makes it perceptually preferable. Solution of this set of normal equations can be found iteratively via Newton's method with the solution of the Wiener-Hopf system of equations as initial conditions. The uniqueness of the solution and the convergence of the algorithm will also be discussed. In section 3.4, some computer experiments will be conducted. The computer simulation results show that compared to the Wiener filter, the optimal IS filter is

superior in terms of spectral matching and output SNR. Finally, the conclusion of this Chapter will be presented in section 3.5.

3.2. Limitations of Realizing the Wiener Filter by Solving the Wiener-Hopf System of Equations

Consider the general digital filter design shown in Figure 3.1. The received signal, $r(n)$, consists of a message signal, $x(n)$, corrupted by a noise sequence, $u(n)$. Assume that the corrupting noise sequence is additive, white, zero-mean, Gaussian distributed with variance σ^2 , and uncorrelated to $x(n)$. Thus, we can write the received signal, $r(n)$, as

$$r(n) = x(n) + u(n) . \quad (3.1)$$

Designing a digital FIR filter is equivalent to estimating a random signal, $x(n)$, on the basis of available observations of related signal, $r(n)$. The resulting estimate, $y(n)$, will be a function of the observation $r(n)$. The goal is to find a set of FIR filter coefficients, a_n , which minimize the linear mean square error, i.e., we want to minimize [Orf90]

$$d = \sum_{n=-\infty}^{\infty} E\{e^2(n)\} \quad (3.2)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} P_e(\omega) d\omega \quad (3.3)$$

where $P_e(\omega)$ is the power spectrum of $e(n)$ defined as

$$\mathbf{e}(\mathbf{n}) = \mathbf{y}(\mathbf{n}) - \mathbf{x}(\mathbf{n}) . \quad (3.4)$$

It is well-known from the orthogonality principle that the coefficients of the optimal MSE FIR filter can also be found by solving [Kay88]

$$\mathbf{P}_{\mathbf{e}\mathbf{r}}(\omega) = \mathbf{0} . \quad (3.5)$$

where $\mathbf{P}_{\mathbf{e}\mathbf{r}}(\omega)$ is the cross spectrum between $\mathbf{e}(\mathbf{n})$ and $\mathbf{r}(\mathbf{n})$, i.e.,

$$\mathbf{P}_{\mathbf{e}\mathbf{r}}(\omega) = \mathbf{E}(\omega)\mathbf{R}^*(\omega) , \quad (3.6)$$

where $\mathbf{E}(\omega)$ and $\mathbf{R}(\omega)$ are the Fourier transform of $\mathbf{e}(\mathbf{n})$ and $\mathbf{r}(\mathbf{n})$ respectively, and $*$ denotes the conjugate operation.

From Figure 3.1 and using equation (3.6), it is easy to show that

$$\mathbf{P}_{\mathbf{e}\mathbf{r}}(\omega) = \mathbf{G}(\omega)[\mathbf{P}_{\mathbf{x}}(\omega) + \sigma^2] - \mathbf{P}_{\mathbf{x}}(\omega) . \quad (3.7)$$

where $\mathbf{P}_{\mathbf{x}}(\omega)$ is power spectrum of $\mathbf{x}(\mathbf{n})$ and $\mathbf{G}(\omega)$ is the Fourier transform of the FIR filter impulse response.

Setting equation (3.7) to zero, we can find $\mathbf{G}_{\text{opt,wiener}}(\omega)$ as

$$\mathbf{G}_{\text{opt,wiener}}(\omega) = \frac{\mathbf{P}_{\mathbf{x}}(\omega)}{\mathbf{P}_{\mathbf{x}}(\omega) + \sigma^2} . \quad (3.8)$$

From equation (3.1), we can find

$$\mathbf{P}_{\mathbf{r}}(\omega) = \mathbf{P}_{\mathbf{x}}(\omega) + \sigma^2 , \quad (3.9)$$

where $\mathbf{P}_{\mathbf{r}}(\omega)$ is the power spectrum of $\mathbf{r}(\mathbf{n})$.

Thus, using equation (3.9) in equation (3.8), we get

$$G_{\text{opt,wienner}}(\omega) = \frac{P_x(\omega)}{P_r(\omega)} . \quad (3.10)$$

Equation (3.10) is known as the Wiener filter frequency response [Orf90].

There are several ways to realize the causal part of equation (3.10). One of the most common ways is to assume that we have a set of N discrete spectrum of both $x(n)$ and $r(n)$, i.e., we have

$$G_{\text{opt,wienner}}(\omega_i) = \frac{P_x(\omega_i)}{P_r(\omega_i)} , \quad 0 \leq i \leq N-1 , \quad (3.11)$$

where $\omega_i = 2\pi/N$ is the i^{th} discrete frequency spanning from 0 to 2π . We are interested in the case where $G(\omega)$ is a linear time-invariant finite impulse response (FIR) filter of order p , i.e.,

$$G(\omega_i) = \sum_{k=0}^p a_k e^{-j\omega_i k} . \quad (3.12)$$

where a_k is the k^{th} coefficient of the FIR filter, $G(\omega)$.

Substituting equation (3.12) into equation (3.11) and rearranging the terms, we get

$$\sum_{k=0}^p a_k e^{-j\omega_i k} P_r(\omega_i) = P_x(\omega_i) . \quad (3.13)$$

Multiplying both sides of equation (3.13) by $1/N$ and taking the inverse discrete fourier transform (IDFT), we finally get

$$\sum_{k=0}^p a_k R_{rr}(i-k) = R_{xx}(i) , 0 \leq i \leq N-1 . \quad (3.14)$$

where the autocorrelation function, $R(i)$, can be defined as [Kay88]

$$R(i) = \frac{1}{N} \sum_{k=0}^{N-1} P(\omega_k) e^{j\omega_k i} . \quad (3.15)$$

Equation (3.14) is the famous Wiener-Hopf system of equations. Note that solving equation (3.14) is equivalent to minimizing the discrete version of equation (3.3), i.e., we are minimizing

$$\mathbf{d} = \frac{1}{N} \sum_{m=0}^{N-1} P_e(\omega_m) . \quad (3.16)$$

Recall that both $x(n)$ and $u(n)$ are discrete signals. As a result, both $P_x(\omega)$ and $P_u(\omega)$, the power spectrum of $u(n)$ which is equal to constant σ^2 , are continuous with the same period of 2π [Opp89]. Assume that the continuous time signals, $x(t)$ and/or $u(t)$, are not bandlimited. Thus, either $P_x(\omega)$ or $P_u(\omega)$ or both are aliased, and so is $P_r(\omega)$. However, from equation (3.12), there is no restriction on $G(\omega)$ to be aliased if $P_r(\omega)$ is. Thus, from equation (3.11), we can see that this process is equivalent to matching the aliased version of the Wiener filter frequency response with the nonaliased FIR filter frequency response.

However, the most undesirable characteristic of the Wiener filter in high quality speech processing is probably that the Wiener filter does not possess the autocorrelation function matching property. To show this, since $G(\omega)$ is a FIR filter, we can write the relationship between processed signal, $y(n)$, and the received signal,

$r(n)$, as

$$y(n) = \sum_{k=0}^p a_k r(n-k) . \quad (3.17)$$

Multiplying equation (3.17) by $y(n-i)$ and taking the expected value on both sides, we get

$$R_{yy}(i) = \sum_{k=0}^p a_k E\{r(n-k)y(n-i)\} \quad (3.18)$$

where the autocorrelation function of $y(n)$, $R_{yy}(i)$, can also be defined as

$$R_{yy}(i) = E\{y(n)y(n-i)\} . \quad (3.19)$$

Using equation (3.17) in equation (3.18), we can show that

$$R_{yy}(i) = \sum_{k=0}^p \sum_{l=0}^p a_k a_l R_{rr}(i+1-k) . \quad (3.20)$$

Comparing equation (3.20) with equation (3.14), we can conclude that

$$R_{yy}(i) \neq R_{xx}(i) , 0 \leq i \leq p . \quad (3.21)$$

Thus, the Wiener filter does not possess the autocorrelation function matching property. In the following section, we will derive a new optimal filter called an optimal IS filter that minimizes the IS distance $x(n)$ and $y(n)$. We will show that this is equivalent to matching the aliased version of the Wiener filter frequency response with the magnitude square of the FIR filter frequency response aliased in the same manner. Furthermore, we also show that this optimal IS filter does satisfy the autocorrelation function matching property whereas the Wiener filter does not.

3.3 Derivation of the Optimal IS Filter

The backward IS distance measure between $x(n)$ and $y(n)$ can be defined as

$$d_{IS} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P_y(\omega)}{P_x(\omega)} - \ln \frac{P_y(\omega)}{P_x(\omega)} - 1 \, d\omega . \quad (3.22)$$

The difference between the forward and backward IS distortion measure is switching the role between the reference signal, $x(n)$, and the processed signal, $y(n)$. Even though, the IS distortion is not symmetric, for a small value of d_{IS} , we can assume that the forward IS distortion measure and the backward IS distortion measure are approximately equal [Gra76].

Before we start deriving the optimal IS filter, we should mention that the beginning of the derivation shares some similarity with El-Jaroudi's works [Elj87, Elj88, Elj89, Elj91]. However, as the derivation goes on, the final results totally differ. This is because in El-Jaroudi, a speech signal is modeled by an AR spectral model which minimizes the forward IS distortion measure between the speech signal and the output of the AR spectral model. In speech modeling, we know that the output of the AR spectral model, the processed signal, is a function only of the AR spectral parameters. However, in our case, we are interested in optimal FIR filter design. From equation (3.17), we can see that the processed signal, $y(n)$, is a function of both FIR filter parameters and the previously received signal, $r(n)$.

From equation (3.1), (3.12), and (3.17), it is easy to show that

$$P_y(\omega) = |G(\omega)|^2 [P_x(\omega) + \sigma^2] \quad (3.23)$$

where $|\bullet|^2$ denotes magnitude square operation. Thus, using equation (3.23) in equation (3.22), we get

$$d_{IS} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P_y(\omega)}{P_x(\omega)} - \ln[P_x(\omega) + \sigma^2] - \ln|G(\omega)|^2 + \ln P_x(\omega) - 1 d\omega . \quad (3.24)$$

We then takes derivative of equation (3.24) respect to a_i and get

$$\frac{\partial d_{IS}}{\partial a_i} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{P_x(\omega)} \frac{\partial P_y(\omega)}{\partial a_i} - \frac{1}{|G(\omega)|^2} \frac{\partial |G(\omega)|^2}{\partial a_i} d\omega . \quad (3.25)$$

From equation (3.23), we can find

$$\frac{\partial P_y(\omega)}{\partial a_i} = [P_x(\omega) + \sigma^2] \frac{\partial |G(\omega)|^2}{\partial a_i} . \quad (3.26)$$

We also know that

$$\frac{\partial |G(\omega)|^2}{\partial a_i} = G^*(\omega) \frac{\partial G(\omega)}{\partial a_i} + G(\omega) \frac{\partial G^*(\omega)}{\partial a_i} . \quad (3.27)$$

Using equation (3.12), equation (3.27) can be simplified to

$$\frac{\partial |G(\omega)|^2}{\partial a_i} = \sum_{k=0}^p a_k e^{j\omega k} e^{-j\omega i} + \sum_{k=0}^p a_k e^{-j\omega k} e^{j\omega i}$$

$$= 2 \sum_{k=0}^P a_k \cos[\omega(k-i)] . \quad (3.28)$$

Inserting equation (3.28), and (3.26) in equation (3.25), we get

$$\begin{aligned} \frac{\partial d_{IS}}{\partial a_i} &= \sum_{k=0}^P a_k \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[\frac{P_x(\omega) + \sigma^2}{P_x(\omega)} - \frac{1}{|G(\omega)|^2} \right] \\ &\quad \times \cos(\omega(k-i)) d\omega . \end{aligned} \quad (3.29)$$

Setting equation (3.29) to zero and rearranging the terms, we get

$$\begin{aligned} &\sum_{k=0}^P a_k \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P_r(\omega)}{P_x(\omega)} \cos[\omega(k-i)] d\omega \\ &= \sum_{k=0}^P a_k \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{|G(\omega)|^2} \cos[\omega(k-i)] d\omega . \end{aligned} \quad (3.30)$$

From equation (3.10), equation (3.30) implies that minimizing the IS distance between $x(n)$ and $y(n)$ is equivalent to matching the aliased version of the Wiener filter response with the magnitude square of frequency response of the FIR filter aliased in the same manner. To realize equation (3.30), we again assume that we have N discrete samples of $P_x(\omega)$. Thus, equation (3.30) reduces to

$$\sum_{k=0}^P a_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_r(\omega_m)}{P_x(\omega_m)} \cos[\omega_m(k-i)]$$

$$= \sum_{k=0}^p a_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{1}{|G(\omega_m)|^2} \cos[\omega_m(k-i)] . \quad (3.31)$$

Note that solving equation (3.31) is equivalent to minimizing the discrete version of equation (3.22), i.e., we are minimizing

$$d_{\text{IS}} = \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_y(\omega_m)}{P_x(\omega_m)} - \ln \frac{P_y(\omega_m)}{P_x(\omega_m)} - 1 . \quad (3.32)$$

We now define

$$H(\omega_m) \triangleq \frac{1}{G(\omega_m)} , \quad m = 0, 1, 2, \dots, N-1 . \quad (3.33)$$

Thus, we have

$$H(\omega_m)G(\omega_m) = 1 . \quad (3.34)$$

Multiply both sides of equation (3.34) by $H^*(\omega)$ and use equation (3.12), we get

$$\sum_{k=0}^p a_k |H(\omega_m)|^2 e^{-j\omega_m k} = H^*(\omega_m) . \quad (3.35)$$

We now define $h(i)$, $0 \leq i \leq N-1$, as the IDFT of $H^*(\omega)$, i.e.,

$$h(i) = \frac{1}{N} \sum_{m=0}^{N-1} H^*(\omega_m) e^{j\omega_m i} , \quad i = 0, \dots, N-1 . \quad (3.36)$$

Using equation (3.35) in equation (3.36), we get

$$h(i) = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{k=0}^p a_k |H^*(\omega_m)|^2 e^{-j\omega_m k} e^{j\omega_m i} ,$$

$$= \sum_{k=0}^p a_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{1}{|G(\omega_m)|^2} \cos[\omega_m(k-i)] . \quad (3.37)$$

since $|G(\omega)|^2$ is an even function.

Using equation (3.37) in equation (3.31), we get

$$\sum_{k=0}^p a_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_r(\omega_m)}{P_x(\omega_m)} \cos[\omega_m(k-i)] = h(i), \quad 0 \leq i \leq p . \quad (3.38)$$

Note that equation (3.38) cannot be solved via a method proposed in [Elj87, Elj88, Elj91] since the output power spectrum is a function of both the input power spectrum and the FIR filter coefficients whereas in El-Jaroudi, the output signal is a function of only the AR parameters. In addition, El-Jaroudi's method requires knowledge of both $P_r(\omega_m)$ and $P_x(\omega_m)$ in advance, which is not practical in many situations. Thus, we propose an alternative technique to approximate the solutions of equation (3.38) from the time domain signal.

We first consider the special case where the FIR filter order, p , is equal to the number of samples in a signal frame, i.e., $p = N-1$. Then, we introduce the intermediate variable, $R_t(i)$, as

$$R_t(i) = \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_r(\omega_m)}{P_x(\omega_m)} \cos(\omega_m i) . \quad (3.39)$$

Note that $R_t(i)$ is in fact the autocorrelation function of the term $P_r(\omega_m)/P_x(\omega_m)$.

Letting $p = N-1$ and using equation (3.39), we can write equation (3.38) as

$$\sum_{k=0}^{N-1} a_k R_t(k-i) = h(i), \quad 0 \leq i \leq N-1. \quad (3.40)$$

Taking IDFT on both sides of equation (3.40) and rearranging the summation term, we get

$$\sum_{k=0}^{N-1} a_k \sum_{i=0}^{N-1} R_t(k-i) e^{j\omega_m i} = H^*(\omega_m). \quad (3.41)$$

From equation (3.39), we can rewrite equation (3.41) as

$$\sum_{k=0}^{N-1} a_k \frac{P_r(\omega_m)}{P_x(\omega_m)} e^{-j\omega_m k} = H^*(\omega_m). \quad (3.42)$$

Using equation (3.12) in equation (3.42), we get

$$G(\omega_m) \frac{P_r(\omega_m)}{P_x(\omega_m)} = H^*(\omega_m). \quad (3.43)$$

Using equation (3.33), equation (3.43) becomes

$$|G(\omega_m)|^2 P_r(\omega_m) = P_x(\omega_m). \quad (3.44)$$

Equation (3.44) implies that minimizing the IS distortion measure between $x(n)$ and $y(n)$ is equivalent to finding the best FIR filter which best matches the power spectrum of $y(n)$ to the power spectrum of $x(n)$. In other words, $G(\omega_m)$ is a linear time-invariant FIR filter that best represents the inverse of adding operation of $P_x(\omega_m)$ and σ^2 . However, we note that we will never fully recover $x(n)$ back since the adding operation is not linear time invariant. For a practical implementation, we want to convert equation (3.44) into the time domain. Hence, using equation (3.12), we can

write equation (3.44) as

$$\sum_{k=0}^{N-1} \sum_{l=0}^{N-1} a_k a_l P_r(\omega_m) e^{j\omega_m(k-l)} = P_x(\omega_m) . \quad (3.45)$$

We now take IDFT of both side of equation (3.45). Thus, we get

$$\sum_{k=0}^{N-1} \sum_{l=0}^{N-1} a_k a_l R_{rr}(i+1-k) = R_{xx}(i) , \quad 0 \leq i \leq N-1 . \quad (3.46)$$

We now make an assumption that $a_i = 0, i > p'$. Thus, equation (3.46) reduces to

$$\sum_{k=0}^{p'} \sum_{l=0}^{p'} a_k a_l R_{rr}(i+1-k) = R_{xx}(i) , \quad 0 \leq i \leq p' . \quad (3.47)$$

Equation (3.47) is our new normal equations. Solving equation (3.47) is much easier than solving equation (3.38) since the autocorrelation functions can be efficiently estimated by the time domain signals.

To solve equation (3.47), we now define

$$f_i = \sum_{k=0}^{p'} \sum_{l=0}^{p'} a_k a_l R_{rr}(i+1-k) - R_{xx}(i) , \quad 0 \leq i \leq p' . \quad (3.48)$$

Taking partial derivative of f_i with respect to a_j , we get

$$\frac{\partial f_i}{\partial a_j} = 2a_j R_{rr}(i) + \sum_{l=0}^{p'} a_l R_{rr}(i-j+1) + \sum_{k=0}^{p'} a_k R_{rr}(i+j-k)$$

$$= \sum_{l=0}^{p'} a_k R_{\pi}(i-j+l) + \sum_{k=0}^{p'} a_k R_{\pi}(i+j-k) . \quad (3.49)$$

We then define a vector A_i , Jacobian matrix, J_i , and function F_i as

$$A_i = [a_0 \ a_1 \ \dots \ a_{p'}]^T , \quad (3.50)$$

$$J_i = \begin{bmatrix} \frac{\partial f_0}{\partial a_0} & \frac{\partial f_0}{\partial a_1} & \dots & \frac{\partial f_0}{\partial a_{p'}} \\ \frac{\partial f_1}{\partial a_0} & \frac{\partial f_1}{\partial a_2} & \dots & \frac{\partial f_1}{\partial a_{p'}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{p'}}{\partial a_0} & \frac{\partial f_{p'}}{\partial a_2} & \dots & \frac{\partial f_{p'}}{\partial a_{p'}} \end{bmatrix} , \quad (3.51)$$

$$F_i = [f_0 \ f_1 \ \dots \ f_{p'}]^T , \quad (3.52)$$

where the subscript i denotes all values are computed at the i^{th} iteration.

Equation (3.47) can be iteratively solved by Newton's method. Newton's algorithm can be written as [Mat87]

$$A_{i+1} = A_i - J_i^{-1} F_i . \quad (3.53)$$

We note that Newton's algorithm is very sensitive to the starting vector, A_0 . If the minimized function is not convex, different starting vectors may lead to different solutions [Rek83]. In addition, even if the minimized function itself is convex, poor selection of the starting vector results in a slow convergence rate. In this problem, we propose that the solution of the Wiener-Hopf system of equations be used as a starting

vector. We will see in the following section that simulation results show that with this choice, Newton's algorithm converges in a very few steps.

To show the existence and uniqueness of the solution, we go back to the relationship between the IS distortion measure and the information discrimination function, $I_N(X,Y)$. We recall from Chapter 2 that the information discrimination function is asymptotically equivalent to the IS distortion measure. Thus, the analysis of the existence and uniqueness of the solution can be approached from the information discrimination function point of view. We start by quoting a theorem from [Gra80], i.e.,

Theorem 3.1: Given two Gaussian distributed random process, $x(n)$ and $z(n)$. Both $x(n)$ and $z(n)$ can be completely characterized by their mean and covariance matrix, μ_x , μ_z , R_x , and R_z , respectively.

We will put constraint on R_y as follow.

a) If there does not exist any positive definite R_y which is an extension of R_z , the $I_N(X,Y)=\infty$. And, hence

$$\min I_N(X,Y) = \infty . \quad (3.54)$$

b) If there exists any positive definite matrix R which is an extension of R_z , then there exists a unique positive definite extension R_y^* called the minimum information discrimination extension of R_z with respect to R_x such that

$$\begin{aligned} \min I_N(X,Y) = & \frac{1}{2} \ln[\det(\mathbf{R}_x)] - \frac{k}{2} + \frac{1}{2} (\mu_z - \mu_x)^T \mathbf{R}_x^{-1} (\mu_z - \mu_x) \\ & + \min_{\mathbf{R}_y \in \mathfrak{S}(\mathbf{R}_z)} \frac{1}{2} \{ \text{tr}(\mathbf{R}_y \mathbf{R}_x^{-1}) - \ln[\det(\mathbf{R}_y)] \} \end{aligned} \quad (3.55)$$

where $\mathfrak{S}(\mathbf{R}_z)$ is the collection of all covariance matrices which are extension of \mathbf{R}_z . It is also shown in [Gra80] that $\mathfrak{S}(\mathbf{R}_z)$ is a convex set of k by k matrices.

To connect the above discussion with our problem, we may view the optimal filtering problem as finding a minimum discrimination function random process, $y(n)$, to represent the random process, $x(n)$. Thus, equation (3.54) and (3.55) imply that for the existence of a solution, it is required that \mathbf{R}_y is positively definite. It is also important to note that even though \mathbf{R}_y^* is unique, the set of the FIR filter coefficients is not. We know from digital signal processing theory that for a FIR filter of order p , there are 2^p sets of the FIR filter coefficients whose roots are complex conjugated to each others, which yield the same power spectrum. Thus, since $\mathfrak{S}(\mathbf{R}_z)$ is convex, if the solution exists, Newton's method will converge to one of the solutions of equation (3.47). However, the recommended set of the FIR filter is the minimum phase one, i.e., all the roots of the FIR coefficients are inside the unit circle since to the minimum phase FIR filter possess several important properties, e.g., minimum phase-lag property, minimum-group delay property, etc. [Opp89]. We should note that once Newton's algorithm converges, the resulting set of the FIR filter coefficients may yield its roots outside the unit circle. In such case, the minimum phase solution can be

found by simply reflecting all the roots which are outside the unit circle back into the unit circle.

In the following section, we will do some computer simulations and discussions. The simulation results show that, compared with the optimal filter obtained from the Wiener-Hopf equations, this new optimal filter not only improves the spectral matching of the estimated power spectrum (smaller IS distortion measure) but also improves the output signal-to-noise (SNR) ratio.

3.4 Computer Simulation Results and Discussions

In this section, we will perform some computer simulations to exhibit the superiority of the new optimal filter over the optimal filter obtained by solving the Wiener-Hopf system of equations. We remind the reader that the goal is to estimate the message signal, $x(n)$, via a filtering technique which minimizes the IS distortion measure between $x(n)$ and $y(n)$. As discussed in the previous section, the optimal IS filter can be obtained by solving equation (3.48) via the Newton's method, equation (3.54) with the solution of the Wiener-Hopf system of equations as the initial condition.

For this example, we define $x(n)$ to be a sequence of the output of an autoregressive process of order 4, AR(4), i.e.,

$$x(n) - 1.352x(n-1) + 1.338x(n-2) - 0.662x(n-3) + 0.24x(n-4) = e(n) . \quad (3.56)$$

where $e(n)$ is a zero-mean white Gaussian noise sequence with variance one.

The message signal, $x(n)$, is corrupted by another zero-mean white Gaussian

noise sequence with variance σ^2 yielding the received signal $r(n)$ as shown in Figure 3.1. For simulation purposes, we will assume that σ^2 is known in advance. The autocorrelation functions of $r(n)$ can be efficiently estimated by [Kay88]

$$\mathbf{R}_r(i) = \frac{1}{N-i} \sum_{k=0}^{N-1} r(k)r(k+i) . \quad (3.57)$$

We note that for a large value of N , equation (3.57) yields a very good estimate of the autocorrelation function of $r(n)$ [Kay88]. For this example, we will use $N=512$.

The simulation starts by solving the Wiener-Hopf system of equations with a given value of filter order, p' , to obtain the initial condition. The optimal IS filter is then found by using Newton's method. The first experiment evaluates the performance of the optimal IS filter with different FIR filter orders. Three different values of p' are selected: 2, 5, and 10. Table 3.1 compares the coefficients of the optimal IS filter with the solutions of the Wiener-Hopf system of equations of the same order in the case where the noise variance is equal to one. The forward and backward IS distortion measures of the optimal IS filter are plotted versus number of iterations with three different values of p' in Figure 3.2 and 3.3. We note that the IS distortion measures of the solution of Wiener-Hopf system of equations is in fact the initial value of the IS distortion measure curves at iteration #0. It is obvious from Figure 3.2 and 3.3 that the optimal IS filter is superior to the solution of the Wiener-Hopf system of equations as we can see sharp drops in all three curves as the iteration starts. Furthermore, as the p' value increases, the IS distortion measure decreases as expected.

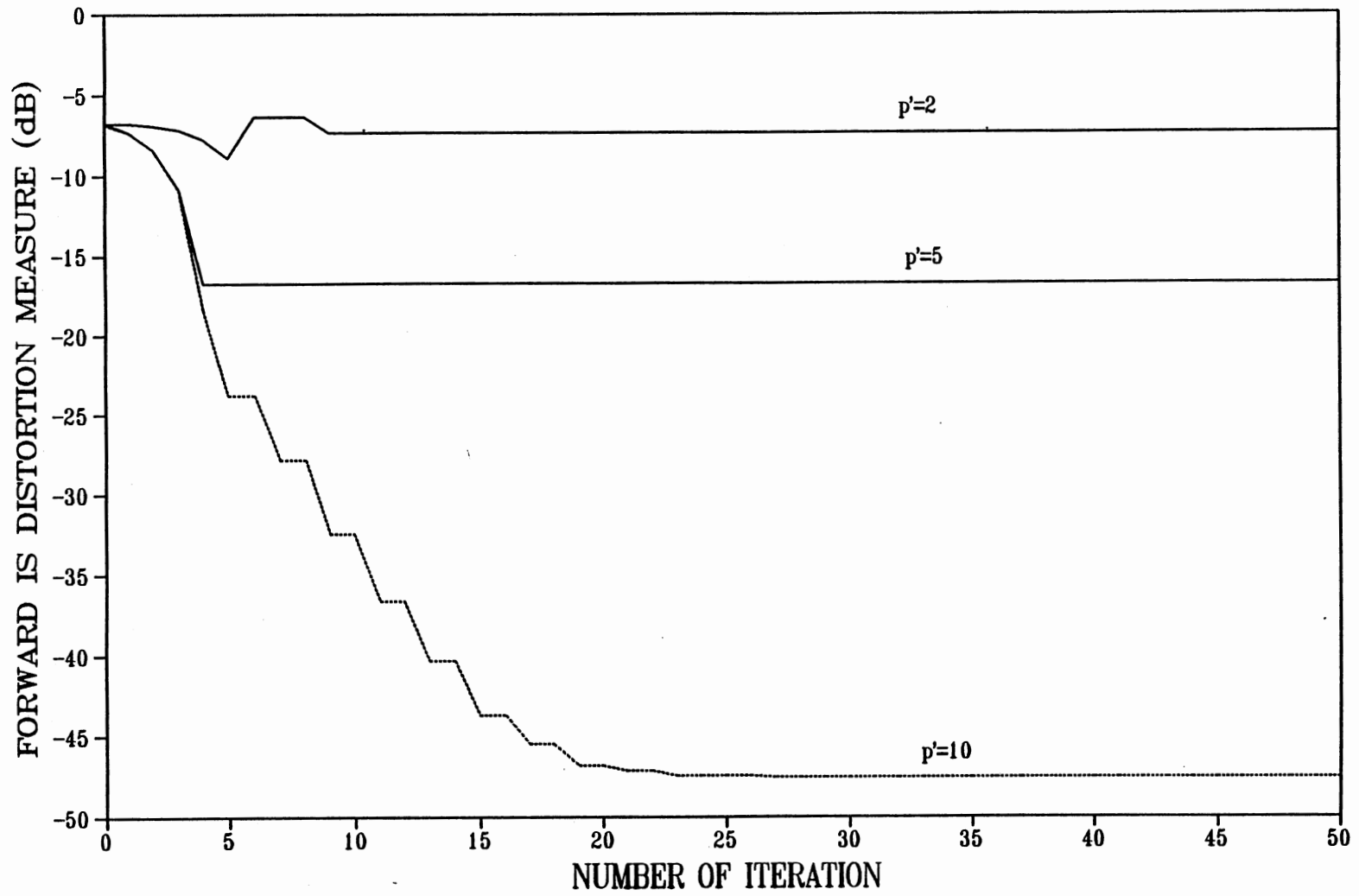


Figure 3.2 Plot of the Forward IS Distortion Measure Versus Number of Iterations (Vary the Filter Order).

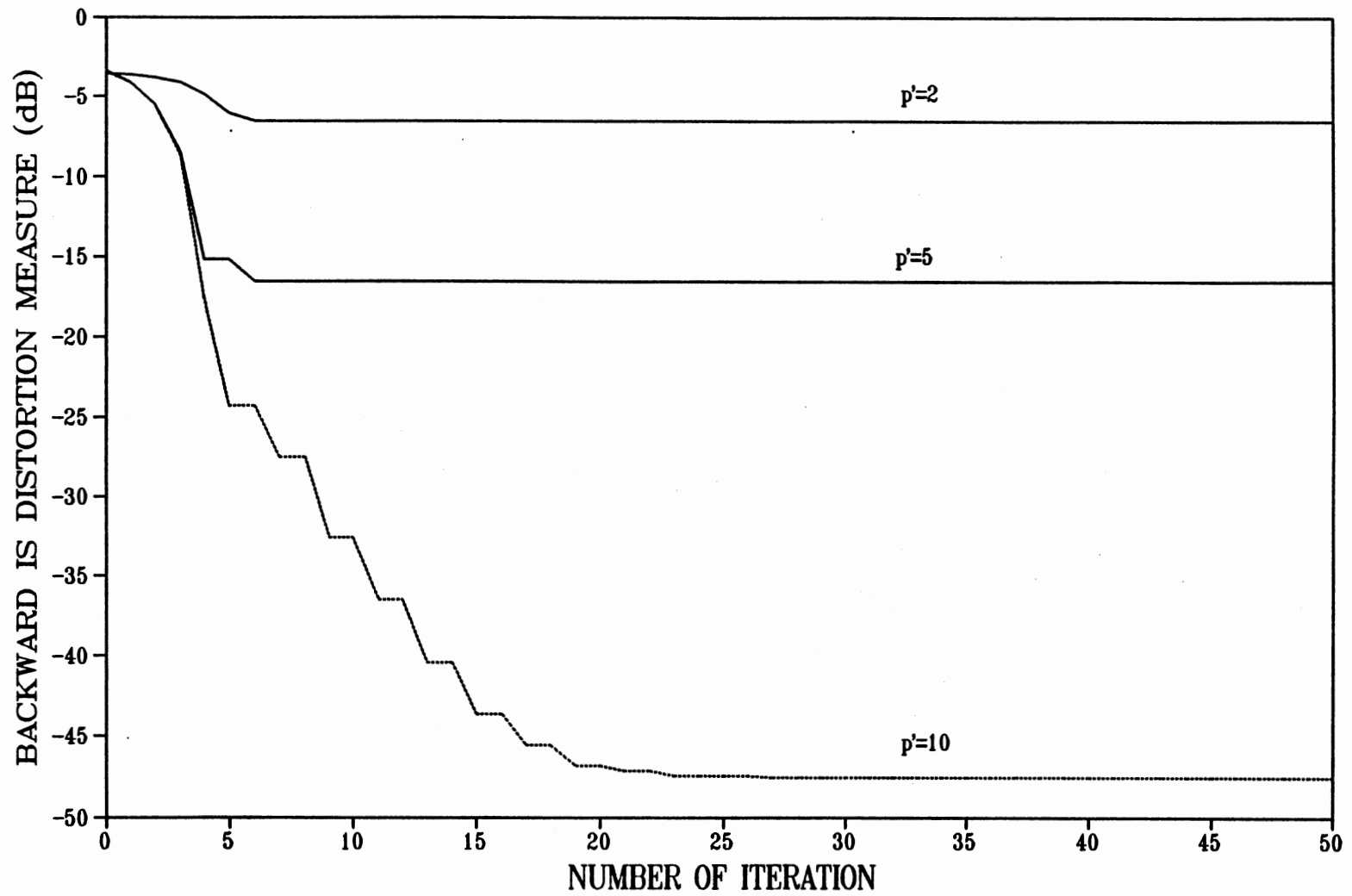


Figure 3.3 Plot of the Backward IS Distortion Measure Versus Number of Iterations (Vary the Filter Order).

TABLE 3.1
COMPARISON OF THE WIENER FILTER COEFFICIENTS
AND THE OPTIMAL IS FILTER COEFFICIENTS

i	Order = 5		Order = 8		Order = 10	
	Wiener	IS	Wiener	IS	Wiener	IS
0	.6773	.5612	.6765	.5681	.6765	.5687
1	.1949	.4321	.1950	.4266	.1950	.4260
2	-.0982	-.0003	-.0974	-.0137	-.0973	-.0137
3	-.0547	-.1317	-.0574	-.0111	-.0573	-.1113
4	.0171	.0210	.0124	-.0081	.0124	-.0082
5	-.0077	.0012	.0023	.0206	.0023	.0224
6			-.0131	-.0019	-.0129	-.0065
7			-.0039	-.0187	-.0033	-.0115
8			.00581	.0092	.0047	.0023
9					.0018	.0043
10					-.0015	-.0007

This is because the assumption $a_i = 0, i > p'$, becomes more and more accurate as p' increases.

In Figure 3.4, we plot output SNR versus number of iteration with three different values of p' . Again, the output SNR of the solution of the Wiener-Hopf system of equations corresponds to the initial value of the SNR curves. From the results, the optimal IS filter is seen to outperform the solution of the Wiener-Hopf system of equations in terms of improving the output SNR. We note that once the iterations converge, there is no significant difference in terms of output SNR for all three curves. However, the difference in the minimum IS distortion measures shown in Figure 3.2 and 3.3 for three different values of p' is quite apparent.

The purpose of the second experiment is to evaluate the performance of the new optimal filter by varying σ^2 . Three different values of σ^2 are also selected: 1, 3, and 5. As in the first experiment, the performance of the optimal filter is evaluated in terms of both forward and backward IS distance measure and output SNR. The results are summarized in Figure 3.5, 3.6, and 3.7. From those Figures, there is no doubt that the optimal IS filter is better than the Wiener filter in terms of both IS distortion measure and output SNR. However, the results show that as σ^2 increase, equivalent to a decrease of the input SNR, the performance of the optimal IS filter slightly degrades in terms of both IS distortion measure and output SNR.

Comparing the forward and backward IS distortion measure results for both experiments, we should note that for small values of distortion measure, both the forward and backward IS distortion measures are approximately equal as expected.

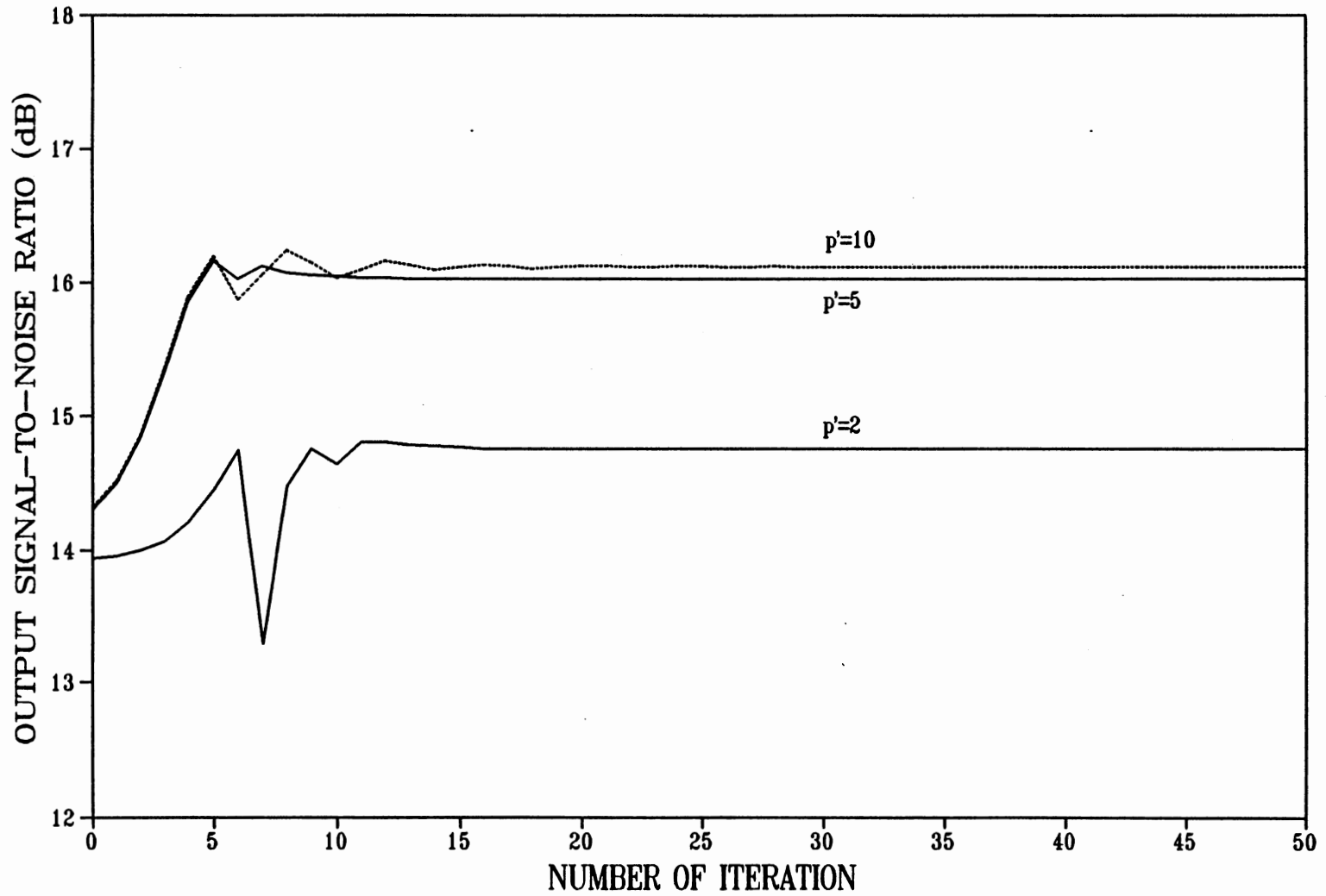


Figure 3.4 Plot of the Output SNR Versus Number of Iterations (Vary the Filter Order).

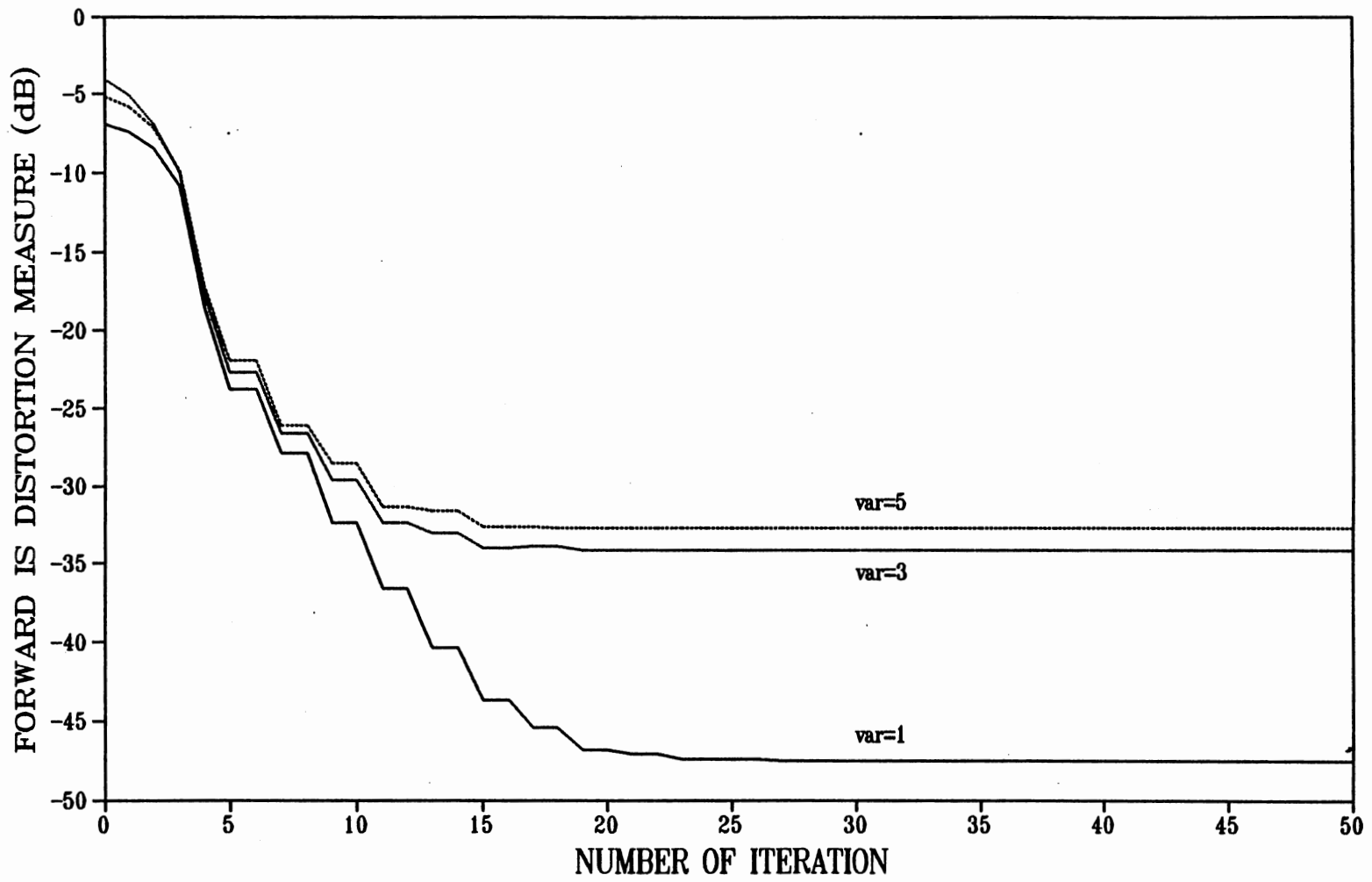


Figure 3.5 Plot of the Forward IS Distortion Measure Versus Number of Iterations (Vary the Corrupting Noise Variance).

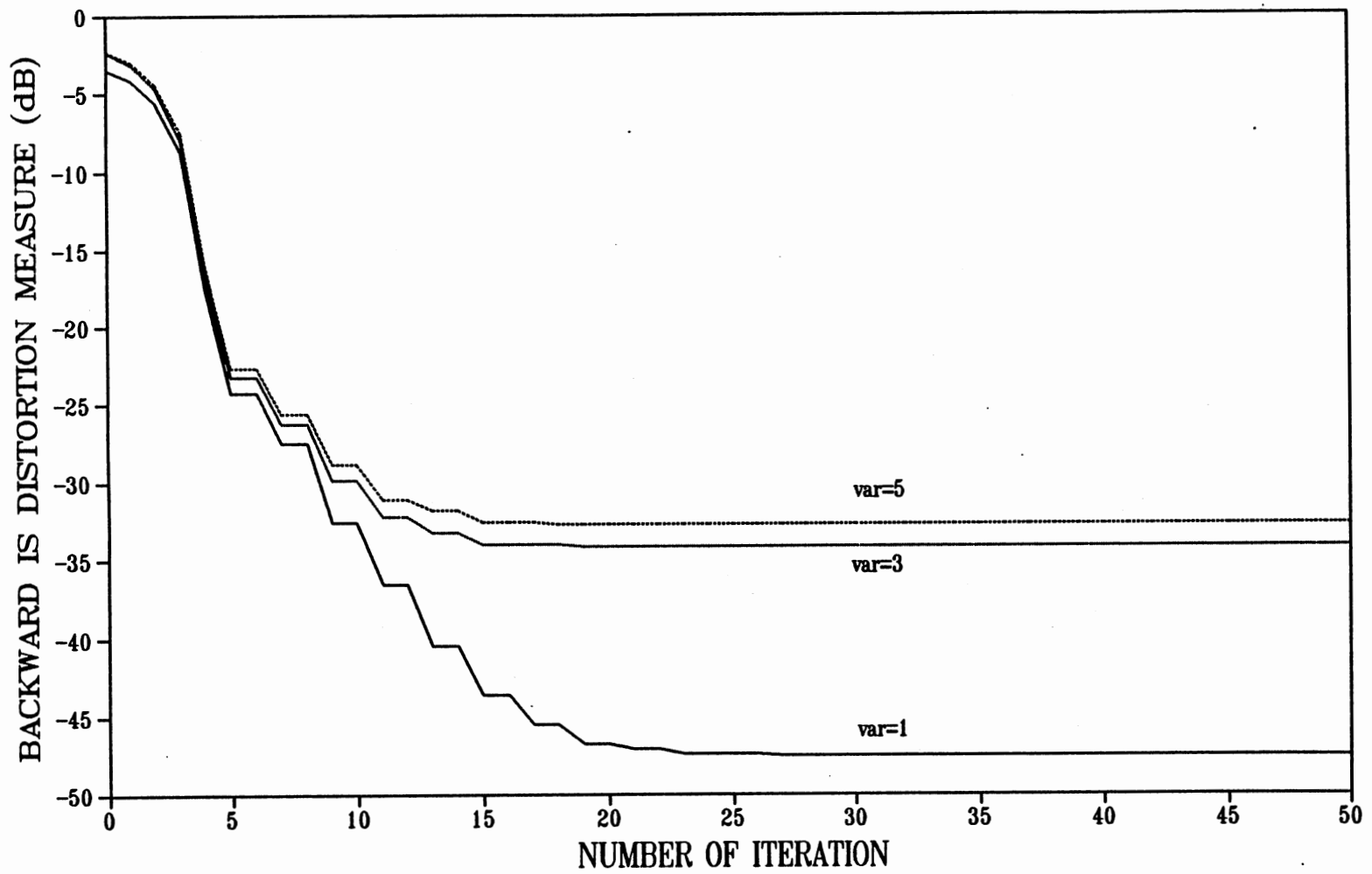


Figure 3.6 Plot of the Backward IS Distortion Measure Versus Number of Iterations (Vary the Corrupting Noise Variance).

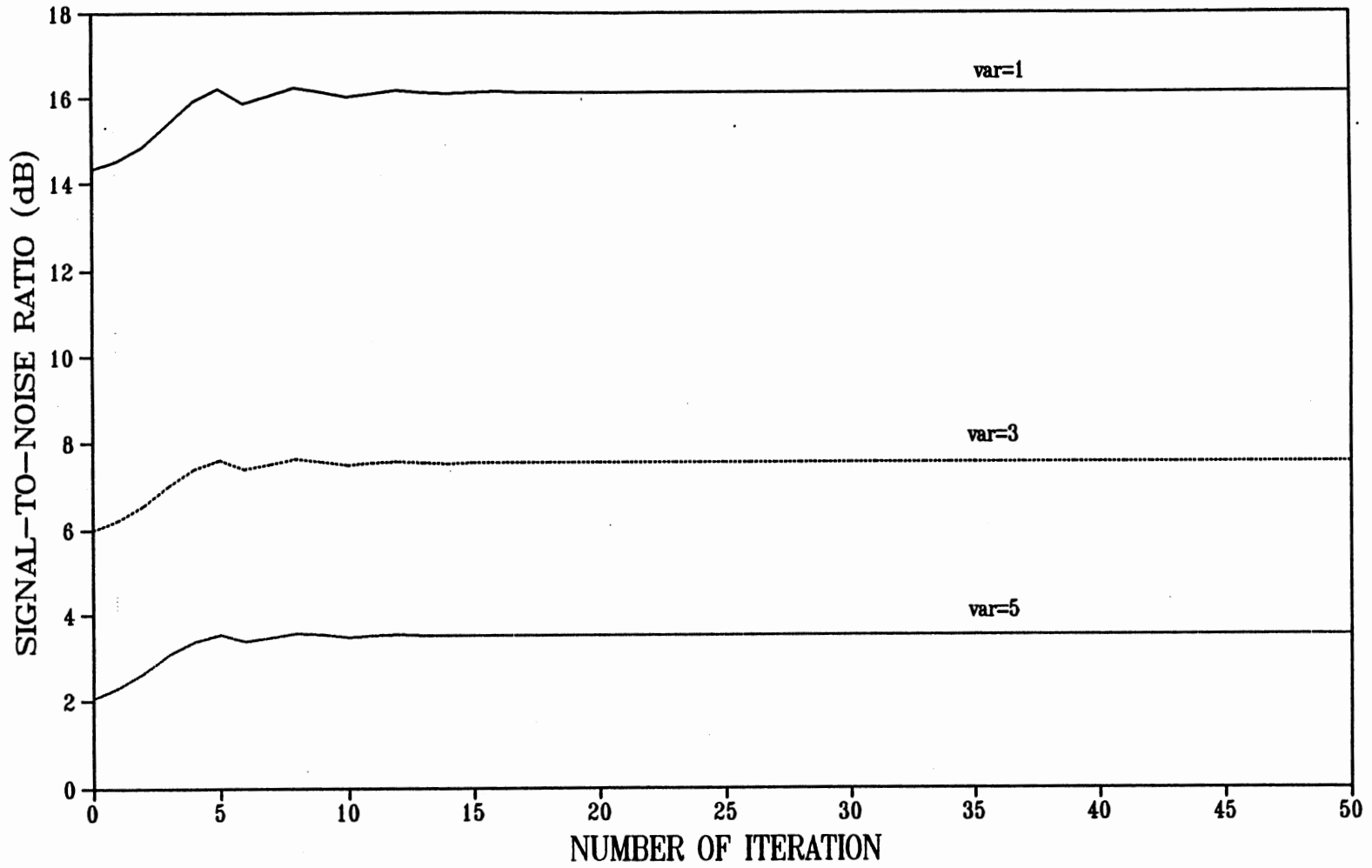


Figure 3.7 Plot of the Output SNR Versus Number of Iterations (Vary the Corrupting Noise Variance).

To illustrate the autocorrelation matching property of the optimal IS filter, we plot the power spectrum of $y(n)$ in the case of $p=5$ and two choices of σ^2 ; 1 and 5, shown in Figure 3.8 and 3.9. For comparison, the power spectrum of $y(n)$ for the case of the solution of the Wiener-Hopf system of equations with same filter order and the same corrupting noise variance are also attached in Figure 3.8 and 3.9. From the results, we can see that the output power spectrum of the optimal IS filter tends to be more robust to increases of the corrupting noise variance than the solution of the Wiener-Hopf system of equations. Thus, we conclude that the optimal IS filter preserves the output spectrum; hence, it preserves the autocorrelation function matching property.

As shown in the previous section, the main reason that the optimal IS filter is superior to the solution of the Wiener-Hopf system of equations is that the optimal IS filter possesses the autocorrelation function matching property. However, another possible reason is because of the property of the IS distortion measure itself. Recall from Chapter 2 that the IS distortion measure is more sensitive to spectral peaks and less to the dips. However, the MSE, which is equivalent to the solution of the Wiener-Hopf equations, weights the contribution of all samples equally. We also learned that the FIR filter tends to have the smoothing property as possessed in the moving average (MA) process. Thus, by minimizing the IS distortion measure, the smoothing effect created by the FIR filter is compensated by the peak sensitivity of the IS distortion measure resulting in better spectral matching.

Another possible advantage that the optimal IS filter may give us is a reduction

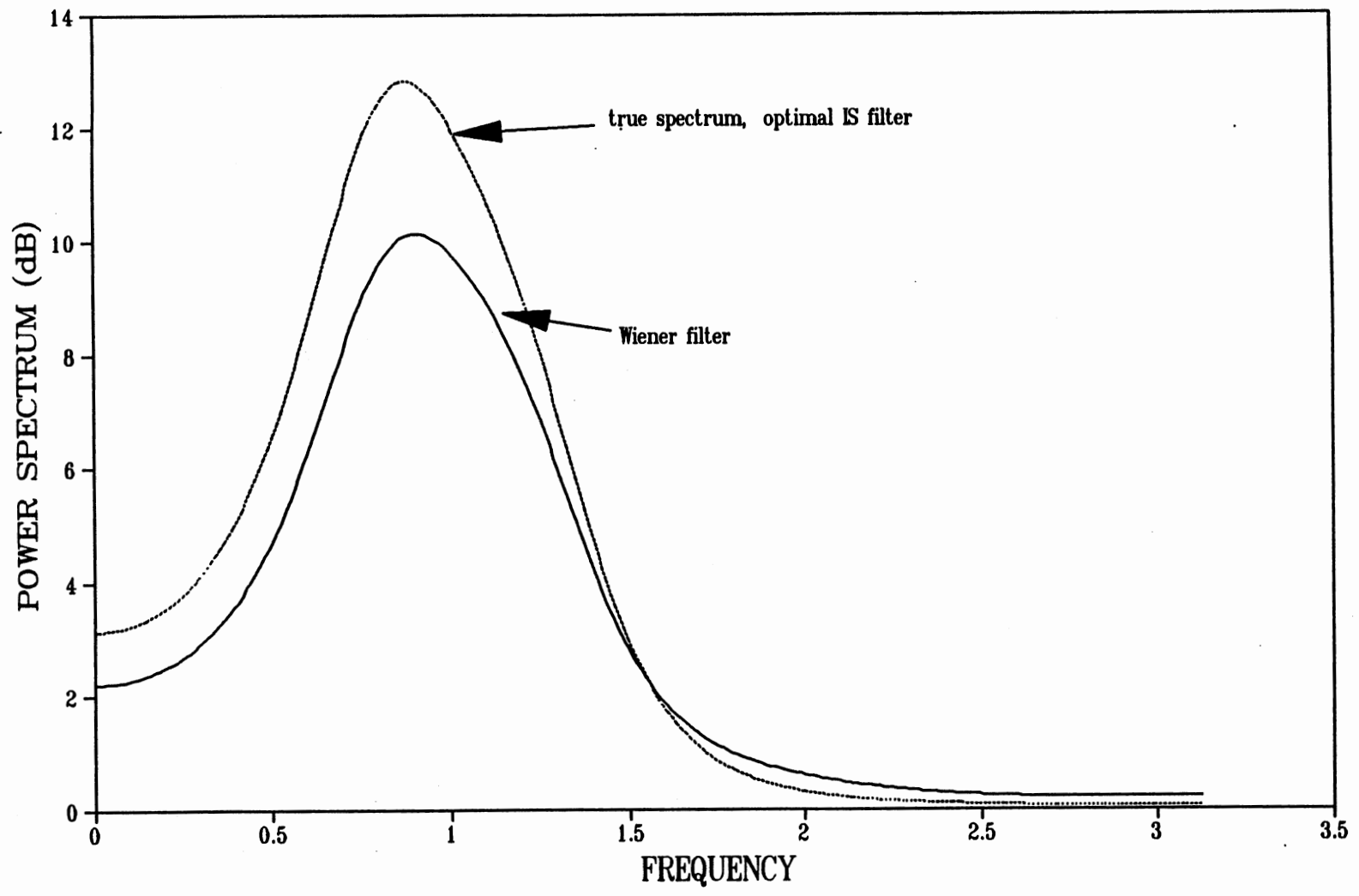


Figure 3.8 Comparison of the Power Spectrums of the Optimal IS Filter Output and the Wiener Filter Output ($\sigma_u^2=1$).

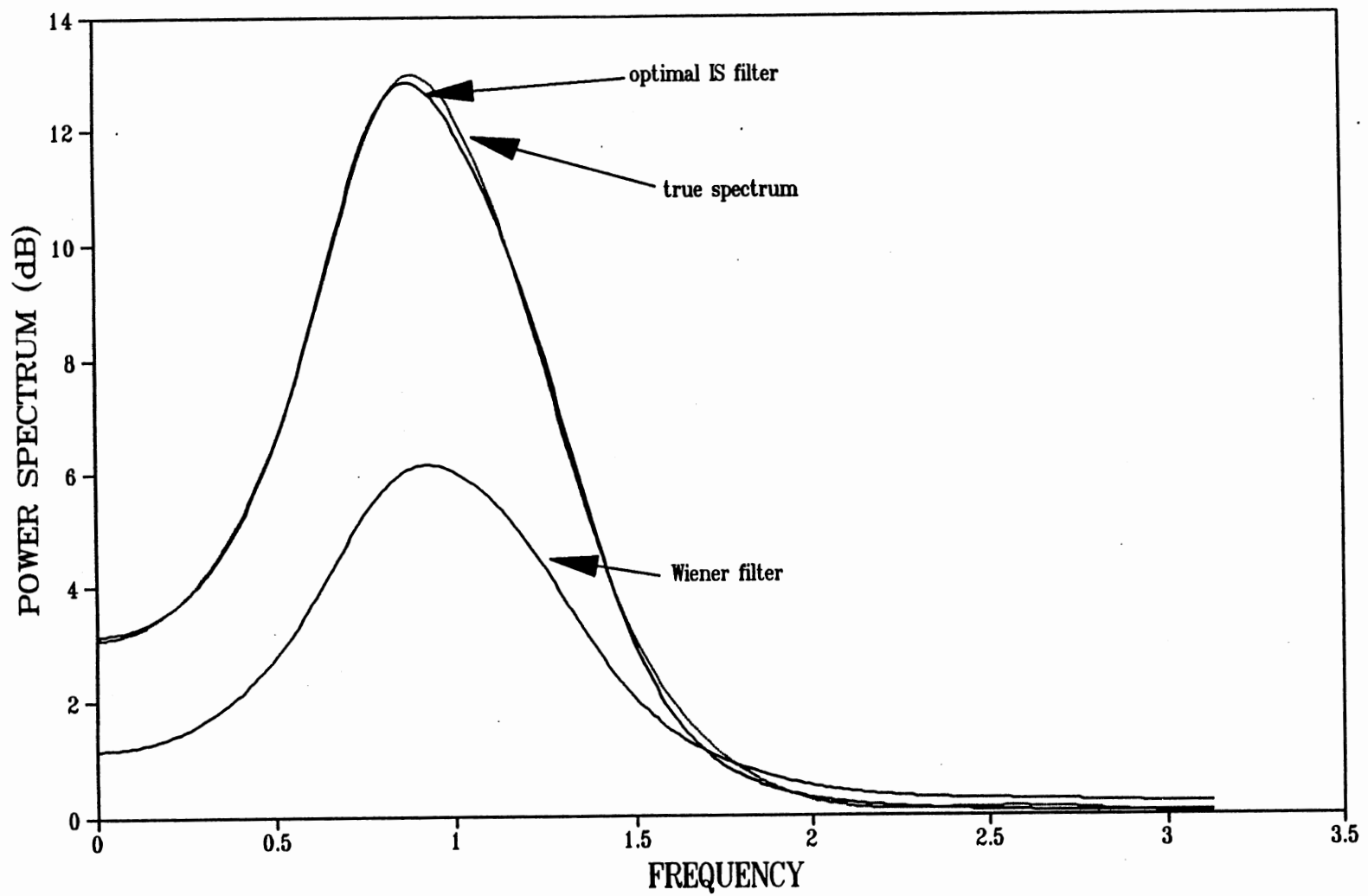


Figure 3.9 Comparison of the Power Spectrums of the Optimal IS Filter Output and the Wiener Filter Output ($\sigma_u^2=5$).

of the required amount of transmitting power at the transmitting end. We know that for a given SNR, the optimal IS filter outperforms the Wiener filter in terms of minimizing the IS distortion measure between the received signal and the unknown transmitted signal (assuming that the noise variance is known in advance). As a result, for a equivalent levels of speech recognition ability, the optimal IS filter communication system should require less transmitting power than a communication system which uses Wiener filters.

3.5. Summary

In this Chapter, we have derived a new optimal FIR filter called the optimal IS filter. This optimal IS filter is obtained by minimizing the IS distortion measure between $x(n)$ and $y(n)$. We also showed that this technique is equivalent to matching the aliased version of the Wiener filter frequency response with the magnitude square of FIR filter frequency response aliased in the same manner. Most importantly, this optimal IS filter possesses the autocorrelation function matching property while the Wiener filter does not. As discussed in Chapter 2, this property results in the optimal IS filter becoming more perceptually desirable. We also performed some computer simulations to illustrate the performance of the optimal IS filter. The simulation results show that the optimal IS filter is superior to the Wiener filter in terms of both spectral matching and the output SNR.

CHAPTER IV

OPTIMAL PRE- AND POST- FILTER DESIGN

4.1 Introduction

In the last Chapter, we proposed a new optimal FIR post-filter called the optimal IS filter. In this Chapter, we propose a method to further improve the performance of the communication system by introducing another FIR filter at the transmitting end of the communication system. It is known that jointly-optimal pre- and post-filter design will enhance the performance of a communication system [Smi65]. This is because the pre-filter transforms the transmitted signal into a form which becomes more robust to the noise in the communication system in some sense. The configuration of pre- and post-filter communication system design is depicted in Figure 4.1. The message signal, $x(n)$, is first corrupted by input noise, $v(n)$. The corrupted signal is then prefiltered by an FIR filter before being transmitted through the communication channel. At the receiver, the received signal, $r(n)$, consists of the transmitted signal corrupted by another noise sequence, called channel noise, $u(n)$. The received signal is then post-filtered again to obtain the estimated signal, $y(n)$.

The design of jointly optimal linear pre- and post-filtering was first pioneered by Costa (without considering the input noise) by minimizing the MSE [Cos52]. The

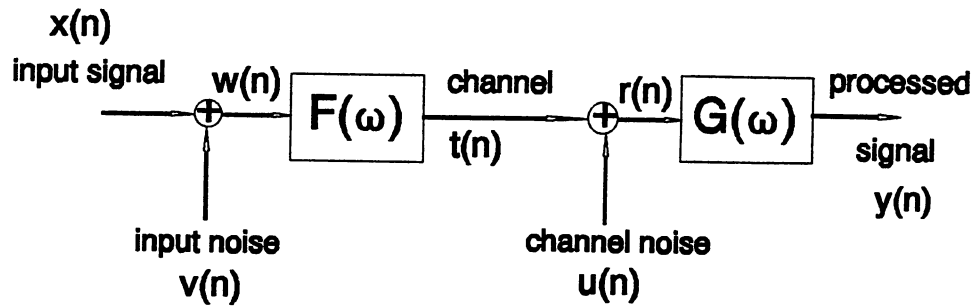


Figure 4.1. Jointly Optimal Pre- and Post-Filter Communication System.

results had been reported to have a moderate or low improvement due to poor choice of parameters. Costa unfortunately elected to express his results in term of (SNR). Twelve years later, Cramer [Cra66] pointed out that the benefit of this technique was in terms of the reduction in the mean square error (MSE), not in terms of improving the SNR [Cra66]. He rederived Costa's work, especially concentrating on the case of the reciprocal filter, where the transfer function of the post-filter is equal to inverse of the transfer function of the pre-filter. In Cramer [Cra66], a significant reduction of the MSE had been reported. Meanwhile, Goodman [Goo66] also applied Costa's results to reduce the quantization noise effect, especially with the application of data storage. With the assumption of using a fine grain quantizer, he showed that the De-emphasis network, i.e., the post-filter, can be asymptotically approximated by the inverse of the

pre-emphasis network, i.e., the pre-filter. Furthermore, the pre-emphasis network behaved like a "half whitening effect" on the input signal.

Brown [Bro61] also considered the joint optimization of pre- and post-filter design for systems that can be modeled by a cascade of pre-filter, sampler, and post-filter when the input consists of signal plus white noise and the absence of channel noise. Chan and Donaldson [Cha71] considered a similar system but with the absence of the input noise and the presence of channel noise. They applied this technique to both pulse code modulation (PCM) and differential pulse code modulation (DPCM) cases for data compression purposes. A considerable amount of bandwidth compression in both cases had been reported in both PCM and DPCM systems.

We note that all of the above literatures deal with a completely analog system (with or without a sampler) under the criteria of minimizing the MSE. Malvar [Mal86, Mal89] was the first person who reconsidered this problem in a completely discrete environment with the presence of both input and channel noises under an input power constraint. The problem is formulated by assuming the sequence $x(n)$, $r(n)$, and $e(n)$ (defined as a difference between the input signal $x(n)$ and the output signal $y(n)$) are cyclostationary [Pap84]. As a result, the MSE can be expressed by the integral of the power spectrum of $e(n)$, obtained by taking the Fourier transform of the autocorrelation function of $e(n)$ averaged over exactly one period. Malvar [Mal86] also showed that the input power constraint is required since the MSE is minimized as $F(\omega)=\infty, \forall\omega$, which is a trivial solution. However, due to the complexity of the problem, Malvar could not come up with closed form solutions for both pre-and

post-filter transfer functions. As a result, the suboptimal pre- and post-filter can be found via the use of a coordinate descent algorithm [Aba82]. Nonetheless, a significant reduction of MSE had been reported.

We note that all of above works are basically done under the criteria of minimizing the MSE. In this Chapter, we will reformulate the problem by designing the joint optimal pre- and post- finite impulse response (FIR) filters that minimize the IS distance measure between $x(n)$ and $y(n)$. In the following section, the new normal equations for optimal pre- and post-filter which minimize the IS distortion measure between $x(n)$ and $y(n)$ will be derived. However, due to the complexity of the problem, as in [Mal86, Mal89], instead of finding the global optimal pre- and post-filter coefficients, we will seek a suboptimal pre- and post-filter solution via the use of Newton's algorithm with the solutions of the Wiener-Hopf equations as the initial condition as in the previous chapter. In section 4.3, some computer simulations are presented and the results are discussed. The simulation results show that these suboptimal filters yield a large improvement in terms of IS distortion measure and output SNR, compared to either the standard Wiener filters or the single optimal post-filter derived in the previous chapter. We also discuss the tradeoff between the joint suboptimal pre- and post filter case, and the single optimal post-filter case. Finally, some conclusions and remarks will be given in section 4.4.

4.2 Derivation of Jointly Optimal Pre- and Post- IS Filters

We now again consider

$$d_{IS} = \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_y(\omega_m)}{P_x(\omega_m)} - \ln \frac{P_y(\omega_m)}{P_x(\omega_m)} - 1 . \quad (4.1)$$

where $P_x(\omega_m)$ and $P_y(\omega_m)$ are power spectrums of $x(n)$ and $y(n)$, respectively. We note that $P_y(\omega_m)$ here is function of both pre- and post-FIR filter coefficients. We assume that $u(n)$ and $v(n)$ are uncorrelated white Gaussian noise with variance, σ_u^2 , σ_v^2 , respectively. Thus, from Figure 4.1, we can show that

$$P_y(\omega_m) = |G(\omega_m)|^2 \{ |F(\omega_m)|^2 [P_x(\omega_m) + \sigma_v^2] + \sigma_u^2 \} . \quad (4.2)$$

where $F(\omega_m)$ and $G(\omega_m)$ are the frequency responses of the FIR pre- and post-filters of order q and p , respectively, which can be defined as

$$F(\omega_m) = \sum_{k=0}^q b_k e^{-j\omega_m k} \quad (4.3)$$

and

$$G(\omega_m) = \sum_{k=0}^p a_k e^{-j\omega_m k} . \quad (4.4)$$

We note that $|\bullet|^2$ denotes the magnitude square operation.

To find the normal equations for optimal pre- and post- FIR filters, we first take partial derivatives of d_{IS} with respect to a_i resulting in

$$\frac{\partial d_{IS}}{\partial a_1} = \frac{1}{N} \sum_{m=0}^{N-1} \left[\frac{1}{P_x(\omega_m)} - \frac{1}{P_y(\omega_m)} \right] \frac{\partial P_y(\omega_m)}{\partial a_1} . \quad (4.5)$$

From equation (4.2), we can find partial derivatives of $P_y(\omega_m)$ with respect to a_1 as

$$\frac{\partial P_y(\omega_m)}{\partial a_1} = P_r(\omega_m) \frac{\partial |G(\omega_m)|^2}{\partial a_1} \quad (4.6)$$

where $P_r(\omega_m)$ is the power spectrum of $r(n)$ equal to

$$P_r(\omega_m) = |F(\omega_m)|^2 [P_x(\omega_m) + \sigma_v^2] + \sigma_u^2 . \quad (4.7)$$

Similar to equation (3.28), we can show that

$$\frac{\partial |G(\omega_m)|^2}{\partial a_1} = 2 \sum_{k=0}^P a_k \cos[\omega_m(k-i)] . \quad (4.8)$$

Using equation (4.8) and equation (4.6) in equation (4.5) yields

$$\begin{aligned} \frac{\partial d_{IS}}{\partial a_1} &= \frac{1}{N} \sum_{m=0}^{N-1} \left[\frac{1}{P_x(\omega_m)} - \frac{1}{P_y(\omega_m)} \right] P_r(\omega_m) \\ &\quad \times 2 \sum_{k=0}^P a_k \cos[\omega_m(k-i)] . \end{aligned} \quad (4.9)$$

From Figure 4.1, we know that

$$P_y(\omega_m) = |G(\omega_m)|^2 P_r(\omega_m) . \quad (4.10)$$

Using equation (4.10) in equation (4.9) and rearranging the terms, we have

$$\frac{\partial d_{IS}}{\partial a_1} = 2 \sum_{k=0}^p a_k \frac{1}{N} \sum_{m=0}^{N-1} \left[\frac{P_r(\omega_m)}{P_x(\omega_m)} - \frac{1}{|G(\omega_m)|^2} \right] \cos[\omega_m(k-i)] . \quad (4.11)$$

Setting equation (4.11) equal to zero, we get

$$\begin{aligned} & \sum_{k=0}^p a_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_r(\omega_m)}{P_x(\omega_m)} \cos[\omega_m(k-i)] \\ &= \sum_{k=0}^p a_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{1}{|G(\omega_m)|^2} \cos[\omega_m(k-i)] . \end{aligned} \quad (4.12)$$

We note that equation (4.12) is very similar to equation (3.31) in Chapter 3 except that $P_r(\omega_m)$ in equation (4.12) is a function of the coefficients of the pre-filter.

To complete the normal equations, we then take the partial derivative of equation (4.1) with respect to b_i . Thus, we get

$$\frac{\partial d_{IS}}{\partial b_i} = \frac{1}{N} \sum_{m=0}^{N-1} \left[\frac{1}{P_x(\omega_m)} - \frac{1}{P_y(\omega_m)} \right] \frac{\partial P_y(\omega_m)}{\partial b_i} . \quad (4.13)$$

From equation (4.2), we can find the partial derivatives of $P_y(\omega_m)$ with respect to b_i as

$$\frac{\partial P_y(\omega_m)}{\partial b_i} = |G(\omega_m)|^2 [P_x(\omega_m) + \sigma_v^2] \frac{\partial |F(\omega_m)|^2}{\partial b_i} . \quad (4.14)$$

In a manner similar to equation (3.28), we can show that

$$\frac{\partial |F(\omega_m)|^2}{\partial b_1} = 2 \sum_{k=0}^q b_k \cos[\omega_m(k-i)] . \quad (4.15)$$

Using equation (4.14) and (4.15), equation (4.13) becomes

$$\begin{aligned} \frac{\partial d_{IS}}{\partial b_1} &= 2 \sum_{k=0}^q b_k \frac{1}{N} \sum_{m=0}^{N-1} \left[\frac{1}{P_x(\omega_m)} - \frac{1}{P_y(\omega_m)} \right] |G(\omega_m)|^2 \\ &\quad \times [P_x(\omega_m) + \sigma_v^2] \cos[\omega_m(k-i)] . \end{aligned} \quad (4.16)$$

Using equation (4.10), equation (4.16) can be simplified to

$$\begin{aligned} \frac{\partial d_{IS}}{\partial b_1} &= 2 \sum_{k=0}^q b_k \frac{1}{N} \sum_{m=0}^{N-1} \left[\frac{|G(\omega_m)|^2}{P_x(\omega_m)} - \frac{1}{P_r(\omega_m)} \right] \\ &\quad \times [P_x(\omega_m) + \sigma_v^2] \cos[\omega_m(k-i)] . \end{aligned} \quad (4.17)$$

Setting equation (4.17) to zero and rearranging the terms, we get

$$\begin{aligned} &\sum_{k=0}^q b_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{|G(\omega_m)|^2}{P_x(\omega_m)} [P_x(\omega_m) + \sigma_v^2] \cos[\omega_m(k-i)] \\ &= \sum_{k=0}^q b_k \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_x(\omega_m) + \sigma_v^2}{P_r(\omega_m)} \cos[\omega_m(k-i)] . \end{aligned} \quad (4.18)$$

Both equation (4.12) and (4.18) constitute a set of normal equations for optimal IS pre- and post- filter design. To solve for the solution of these two normal equations, we first define

$$R_1(i) = \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_r(\omega_m)}{P_x(\omega_m)} \cos \omega_m i, \quad (4.19)$$

$$R_2(i) = \frac{1}{N} \sum_{m=0}^{N-1} \frac{1}{|G(\omega_m)|^2} \cos \omega_m i, \quad (4.20)$$

$$R_3(i) = \frac{1}{N} \sum_{m=0}^{N-1} \frac{|G(\omega_m)|^2}{P_x(\omega_m)} [P_x(\omega_m) + \sigma_v^2] \cos[\omega_m i], \quad (4.21)$$

and

$$R_4(i) = \frac{1}{N} \sum_{m=0}^{N-1} \frac{P_x(\omega_m) + \sigma_v^2}{P_r(\omega_m)} \cos[\omega_m i]. \quad (4.22)$$

Thus, we can rewrite equation (4.12) as

$$\sum_{k=0}^p a_k R_1(k-i) = \sum_{k=0}^p a_k R_2(k-i). \quad (4.23)$$

Recall that convolution in the time domain is equivalent to multiplication in the frequency domain. As in Chapter 3, for a sufficiently large p and q that $a_i = 0$, $i > p$, and $b_j = 0$, $j > q$, taking the DFT of both sides of equation (4.23) can be approximated by

$$G(\omega_m) \frac{P_r(\omega_m)}{P_x(\omega_m)} = G(\omega_m) \frac{1}{|G(\omega_m)|^2},$$

or

$$|G(\omega_m)|^2 P_r(\omega_m) = P_x(\omega_m) . \quad (4.24)$$

Using equation (4.21) and (4.22), we can rewrite equation (4.17) as

$$\sum_{k=0}^q b_k R_3(k-i) = \sum_{k=0}^q b_k R_4(k-i) . \quad (4.25)$$

As before, for a sufficiently large p and q , we can show that

$$\frac{|G(\omega_m)|^2}{P_x(\omega_m)} [P_x(\omega_m) + \sigma_v^2] = \frac{P_x(\omega_m) + \sigma_v^2}{P_r(\omega_m)} ,$$

or

$$|G(\omega_m)|^2 P_r(\omega_m) = P_x(\omega_m) .$$

Note that the above equation is exactly the same as equation (4.24). The time-domain representation of equation (4.24) can be written as

$$\sum_{k=0}^p \sum_{l=0}^p a_k a_l R_{rr}(i-k+l) = R_{xx}(i) . \quad (4.26)$$

We note the autocorrelation function of $r(n)$, $R_{rr}(i)$, is a function of the pre-filter coefficients. Using equation (3.20) in Chapter 3, we can see that the left-handed side of equation (4.26) is in fact equal to the autocorrelation function of $y(n)$. Thus, our optimal pre- and post-filters preserve the autocorrelation function matching property.

As in Chapter 3, the suboptimal solution of equation (4.26) can be solved via Newton's method. Thus, we define

$$f_i = \sum_{k=0}^p \sum_{l=0}^p a_k a_l R_{\text{xx}}(i-k+l) - R_{\text{xx}}(i) . \quad (4.27)$$

Solving equation (4.26) is therefore equivalent to solving

$$f_i = 0, \quad 0 \leq i \leq p+q+1 . \quad (4.28)$$

For Newton's algorithm, we define

$$\Theta_j = [a_0 \ a_1 \ \dots \ a_p \ b_0 \ b_1 \ \dots \ b_q]^T , \quad (4.29)$$

$$F_j = [f_0 \ f_1 \ \dots \ f_{p+q+1}]^T , \quad (4.30)$$

and

$$J_j = \begin{bmatrix} \frac{\partial f_0}{\partial a_0} & \dots & \frac{\partial f_0}{\partial a_p} & \frac{\partial f_0}{\partial b_0} & \dots & \frac{\partial f_0}{\partial b_q} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_{p+q+1}}{\partial a_0} & \dots & \frac{\partial f_{p+q+1}}{\partial a_p} & \frac{\partial f_{p+q+1}}{\partial b_0} & \dots & \frac{\partial f_{p+q+1}}{\partial b_q} \end{bmatrix} , \quad (4.31)$$

where the j subscript denotes that all values are obtained at the j^{th} iteration.

Newton's algorithm can be expressed as [Mat87]

$$\Theta_{j+1} = \Theta_j - J_j^{-1} F_j . \quad (4.32)$$

To construct the Jacobian matrix, J_j , we need to find the partial derivative of f_i with respect to a_j and b_j . As shown in equation (3.50) in Chapter 3, we can find

$$\frac{\partial f_1}{\partial a_j} = \sum_{l=0}^p a_l R_r(i-j+1) + \sum_{l=0}^p a_l R_r(i-1+j) . \quad (4.33)$$

where $R_r(i)$ can be efficiently estimated by equation (3.57) in Chapter 3.

Next, we take partial derivative of equation (4.27) with respect to b_j resulting in

$$\frac{\partial f_1}{\partial b_j} = \sum_{k=0}^p \sum_{l=0}^p a_k a_l \frac{\partial R_r(i-k+1)}{\partial b_j} . \quad (4.34)$$

From Figure 4.1, we can see that

$$r(n) = t(n) + u(n) . \quad (4.35)$$

Since $v(n)$ and $u(n)$ are uncorrelated to each other and $x(n)$, we know that

$$R_r(i) = R_u(i) + \sigma_u^2 \delta(i) , \quad (4.36)$$

where $R_u(i)$ is the autocorrelation function of the transmitting signal $t(n)$, and $\delta(i)$ is dirac function defined as

$$\delta(i) = \begin{cases} 1, & i=0 \\ 0, & \text{otherwise} \end{cases} . \quad (4.37)$$

Using equation (4.36), equation (4.34) becomes

$$\frac{\partial f_1}{\partial b_j} = \sum_{k=0}^p \sum_{l=0}^p a_k a_l \frac{\partial R_u(i-k+1)}{\partial b_j} . \quad (4.38)$$

From Figure 4.1, as in equation (3.20), we can show that

$$\mathbf{R}_u(i) = \sum_{k=0}^q \sum_{l=0}^q b_k b_l \mathbf{R}_{ww}(i-k+l) . \quad (4.39)$$

where $\mathbf{R}_{ww}(i)$ is the autocorrelation of $w(n)$, defined as a sum of $x(n)$ and $v(n)$, which can be defined as

$$\mathbf{R}_{ww}(i) = \mathbf{R}_{xx}(i) + \sigma_v^2 \delta(i) , \quad (4.40)$$

since $v(n)$ is uncorrelated to $x(n)$.

As in equation (3.49) in Chapter 2, we can show that

$$\frac{\partial \mathbf{R}_u(i)}{\partial b_j} = \sum_{l=0}^q b_l \mathbf{R}_{ww}(i-j+l) + \sum_{l=0}^q b_l \mathbf{R}_{ww}(i-1+j) . \quad (4.41)$$

In addition, we recall that

$$\frac{\partial \mathbf{R}_u(-i)}{\partial b_j} = \frac{\partial \mathbf{R}_u(i)}{\partial b_j} . \quad (4.42)$$

Thus, using equation (4.42) and (4.41), we can construct equation (4.38).

We note again that Newton's algorithm is sensitive to the starting vector. Due to the complex nonlinearity of this problem, this algorithm may converge on a suboptimal solution. There is no guarantee that the resulting solution will be the global optimal solution. Thus, a good (bad) choice of the starting vector will lead to good (bad) suboptimal solution. We use the solution of the single optimal IS post-filter found in Chapter 3 as the starting vector. Even though the resulting solution may not be the global one, it should at least guarantee that our jointly optimal design will perform equally or better than the result found in Chapter 3. In the following

section, we will perform some computer simulations to demonstrate the performance of the jointly suboptimal pre- and post- filter design. Simulation results show that as expected, the jointly optimal pre- and post-filter design provides better matching than the single optimal IS post-filter in Chapter 3. In addition, with the solution of the single optimal IS post-filter as the starting vector the algorithm converges in a few steps.

4.3 Computer Simulation Results and Discussions

In this section, we perform some computer simulations to exhibit the performance of the jointly-suboptimal pre- and post-filters. Again, for this example, we consider the same AR(4) sequence used in Chapter 3 (equation 3.56). The first experiment examines the effect of varying both pre- and post-filter orders. We assume that both σ_v^2 and σ_u^2 are known in advance or can be accurately estimated somehow, e.g., from the link equation. For the first experiment, we let $\sigma_v^2 = \sigma_u^2 = 1.0$. We also assume that both pre- and post-filter have the same order, i.e., $p = q$. The FIR filter order varies from 2, 4, and 5. The simulation results are displayed in terms of the IS distortion measure and SNR versus the number of iterations as illustrated in Figure 4.2 and Figure 4.3. From Figure 4.2 and 4.3, we can see that the performance of the jointly optimal system improves as the FIR filter order increases as is noted from the significant reduction of the IS distortion measure and increase of the output SNR. In addition, we note that the algorithm converges within a few iteration steps. This implies that suboptimal solutions are within the vicinity of the solution of the single

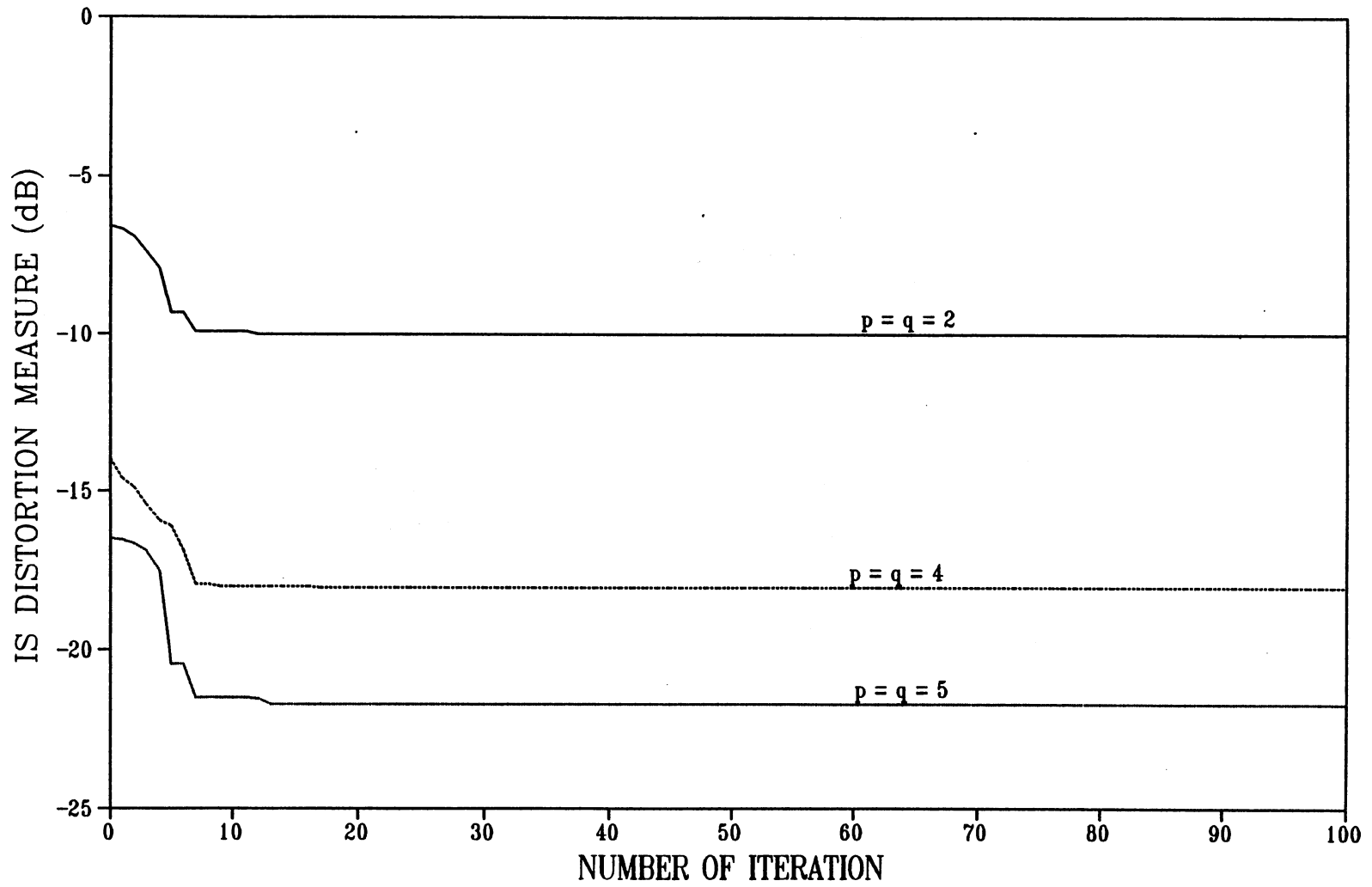


Figure 4.2 Plot of the IS Distortion Measure Versus Number of Iterations (Vary the Filter Orders).

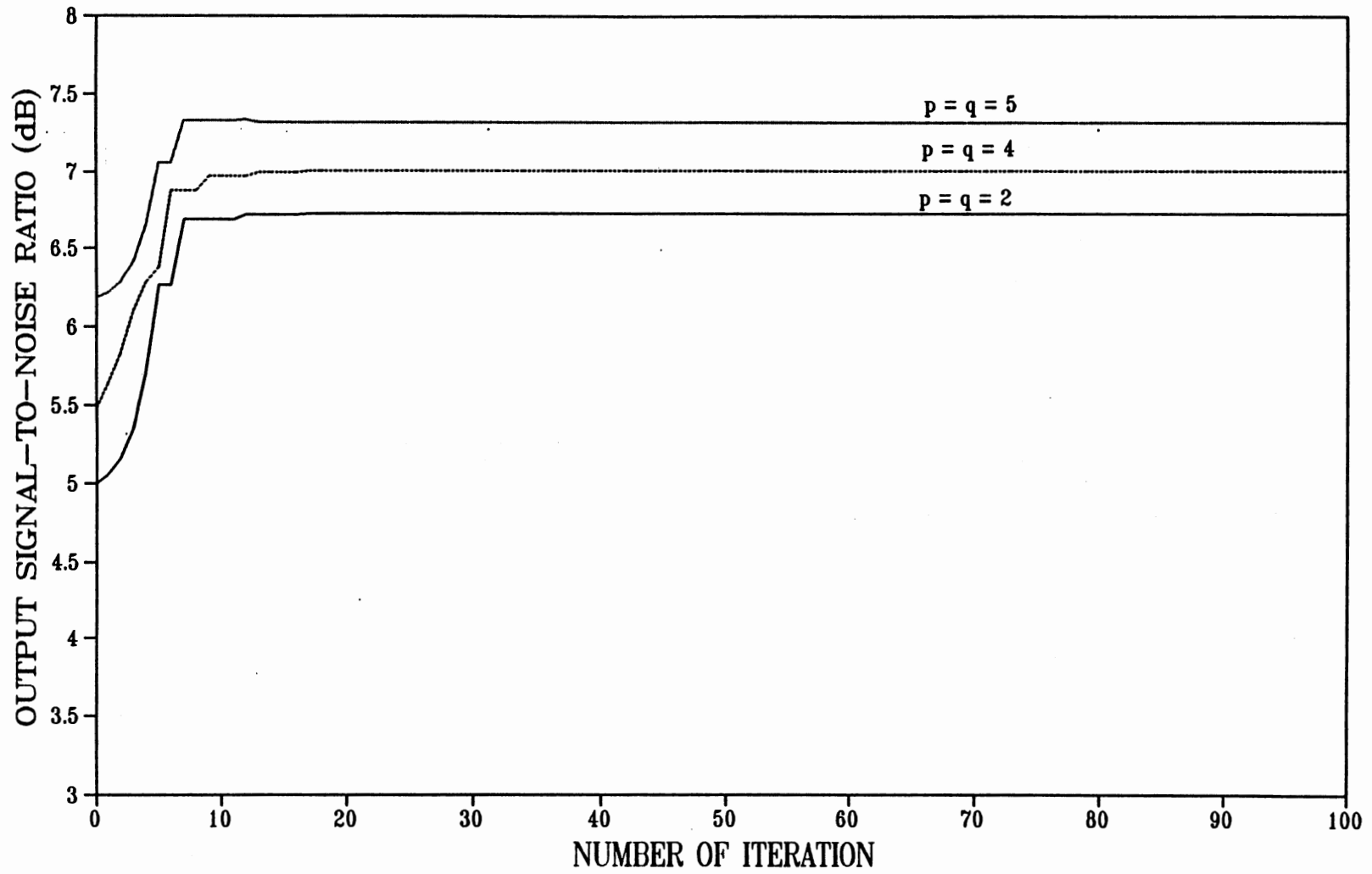


Figure 4.3 Plot of the Output SNR Versus Number of Iterations (Vary the Filter Orders).

optimal IS post-filter. Thus, selection of the solution of the single optimal IS post-filter as the initial condition is a good choice.

The second experiment examines the effect of both input noise variance and channel noise variance. Figure 4.4 and 4.5 show the plots of the jointly suboptimal pre- and post-filter performance under three different value of σ_v^2 and σ_u^2 . From both results, as expected, the performance of the jointly suboptimal filters degrades as the noise variances increases.

One may have questions as to how the input noise and the channel noise effect the performance of the jointly suboptimal filters. This question is addressed into the next experiment.

In the third experiment we vary σ_v^2 and σ_u^2 in such a way that $\sigma_v^2 + \sigma_u^2 = 3$. The performance of the jointly suboptimal filter under three different combination of σ_v^2 and σ_u^2 are shown in Figure 4.6 and 4.7 in terms of both IS distortion measure and SNR. We can see that the case where σ_v^2 is greater than σ_u^2 shows the worst performance while the case where σ_v^2 is less than σ_u^2 shows the best performance. The reason can be explained as follows. The total noise at the input of the suboptimal post-filter consists of the channel noise and the pre-filtered input noise. We also know that the FIR filter does possess the smoothing property as in the MA process. Thus, the spectrum at the input of the post-filter in the case where $\sigma_v^2 > \sigma_u^2$ is smoother than the spectrum in the case where $\sigma_v^2 < \sigma_u^2$, as a result, poorer performance is exhibited in the computer simulations.

To compare the jointly suboptimal filter design with the single optimal IS post-

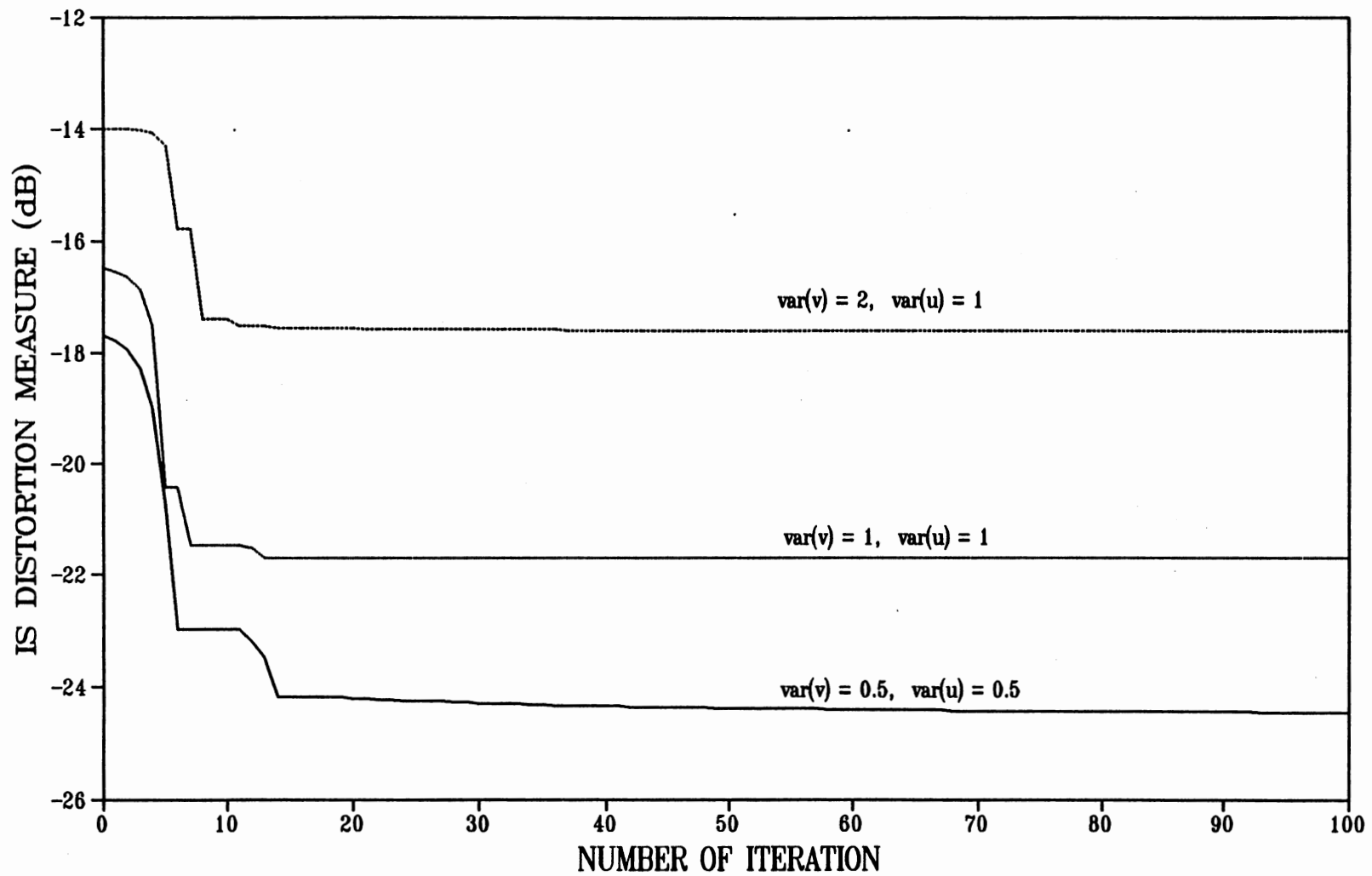


Figure 4.4 Plot of the IS Distortion Measure Versus Number of Iterations (Vary Corrupting Noise Variances).

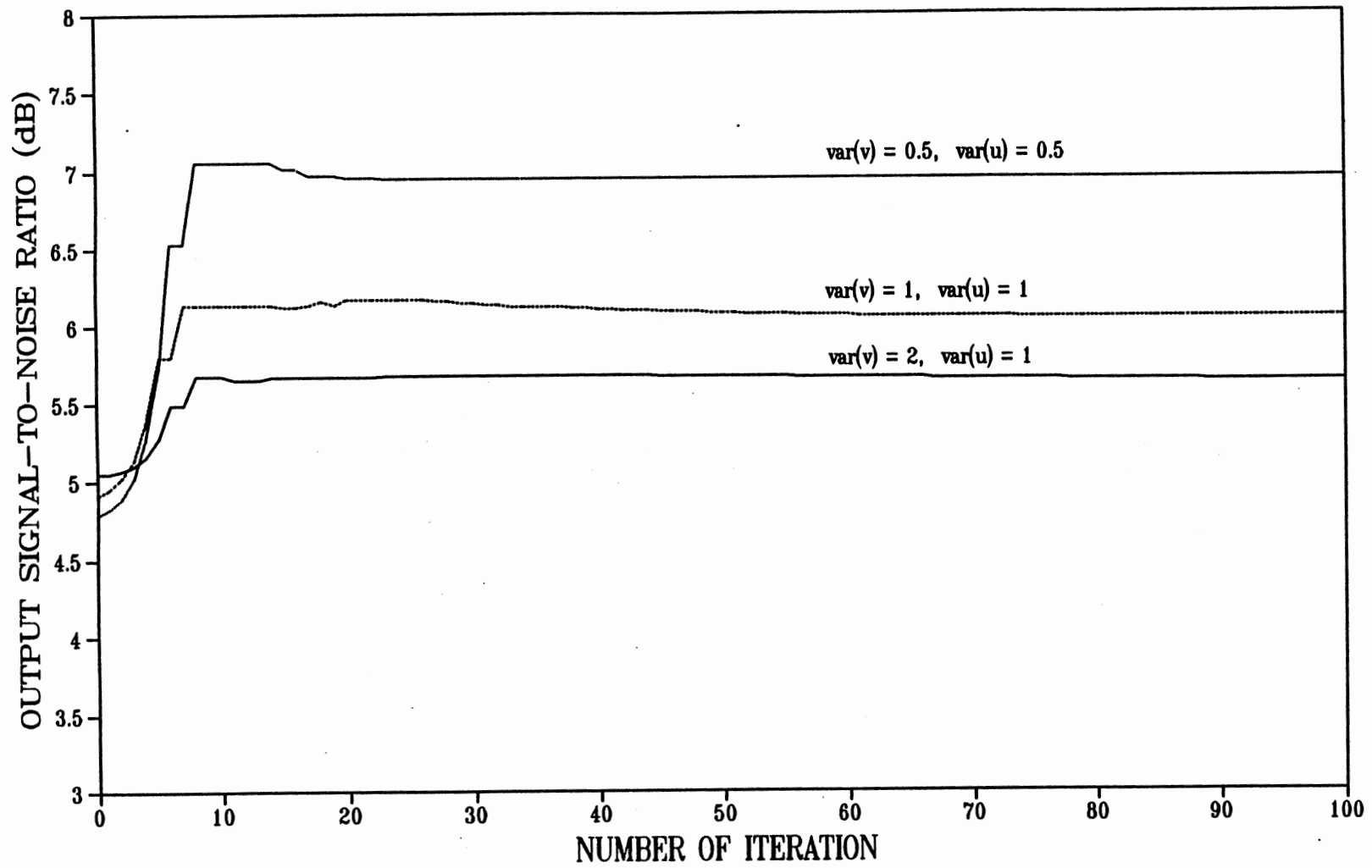


Figure 4.5 Plot of the Output SNR Versus Number of Iterations (vary Corrupting Noise Variances).

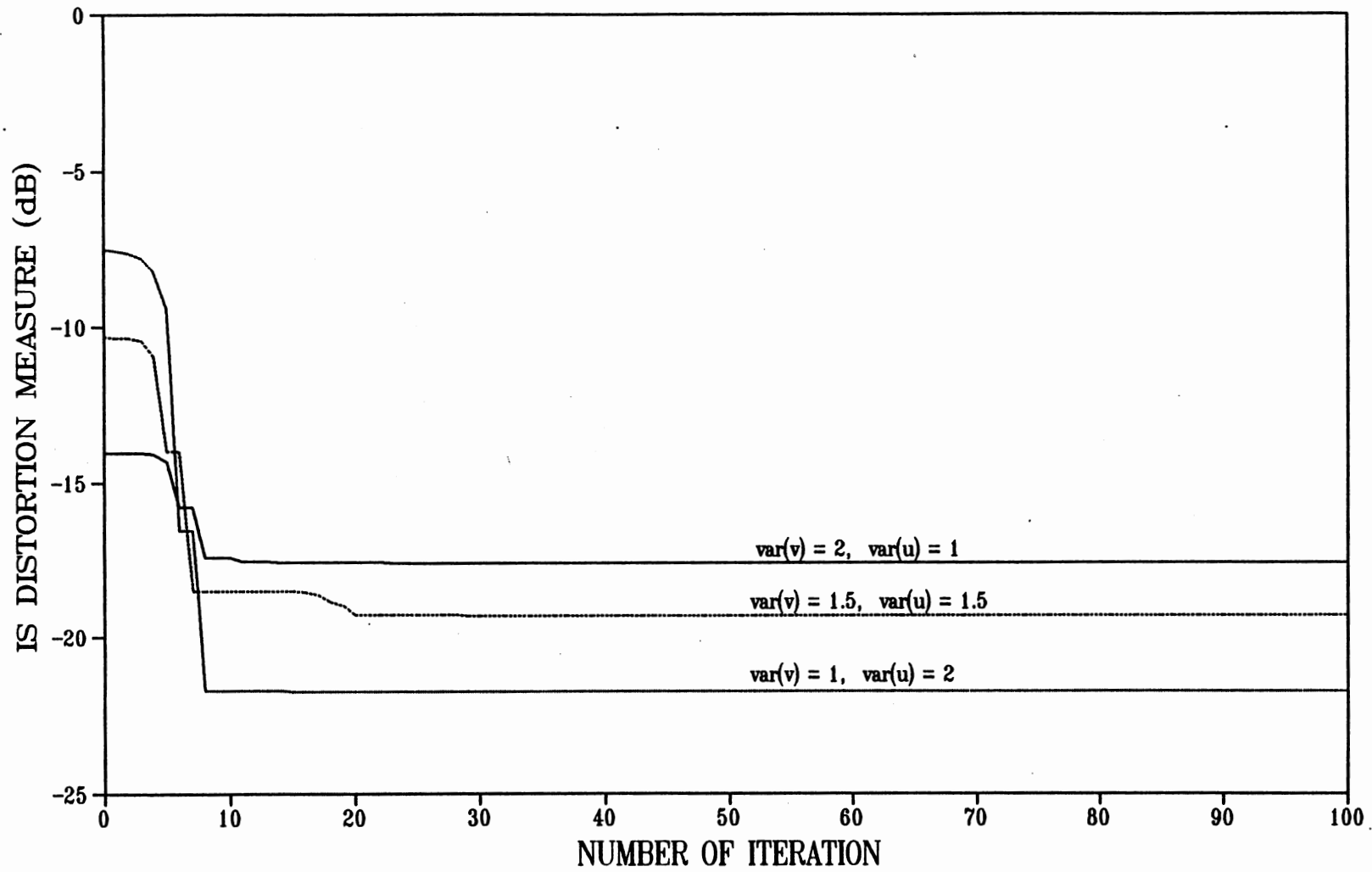


Figure 4.6 Plot of the IS Distortion Measure Versus Number of Iterations for Three Combinations of Corrupting Noise Variances.

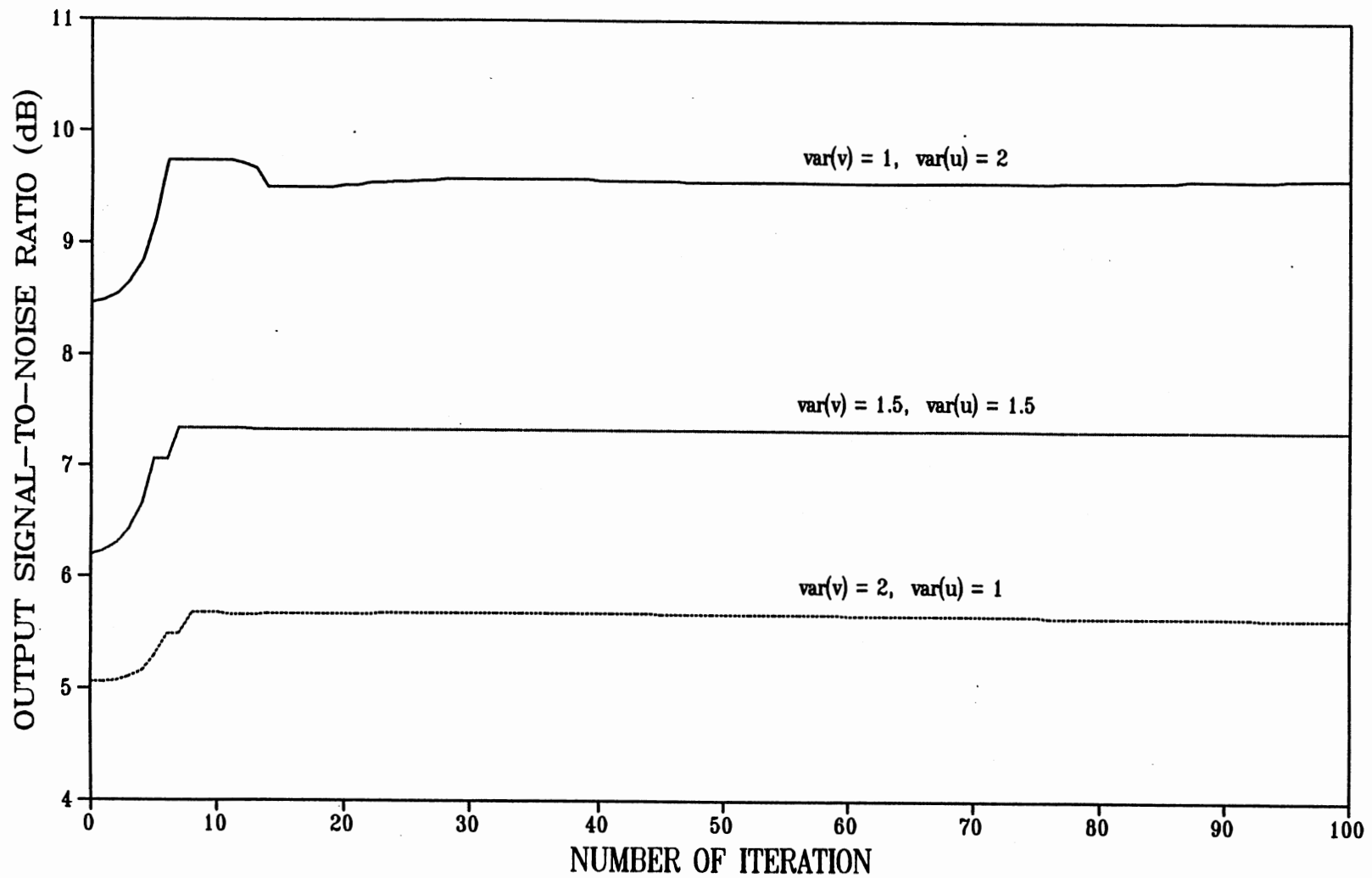


Figure 4.7 Plot of the Output SNR Versus Number of Iterations for Three Different Combinations of Corrupting Noise Variances.

filter design, we summarize the optimal values of the IS distortion measure and the output SNR under different value of noise variance in Table 4.1. From table 4.1, there is no doubt that the jointly suboptimal design is superior than the single optimal IS post-filter design in terms of IS distortion measure.

TABLE 4.1
COMPARISON OF THE IS DISTORTION MEASURE OF THE
SINGLE OPTIMAL IS FILTER AND JOINTLY
SUBOPTIMAL SYSTEM

$\sigma_v^2 + \sigma_u^2$	1	2
single optimal IS filter*	3.19e-02	1.14e-02
jointly suboptimal system	3.58e-03	6.63e-03

note: * $\sigma_v^2 = 0$ for single optimal IS filter.

4.4 Summary

So far, we have shown that an improvement in term of spectral matching can be made by using jointly optimal pre- and post-filter design. The use of the pre-filter before transmitting a signal through a communication channel will change the transmitting signal to a more robust form compared with the existing noise in the communication system. In this Chapter, we derived the normal equations of jointly optimal pre- and post-filter design. We also showed that even though the system complexity increased, the jointly optimal pre- and post-filter design still preserved the autocorrelation function matching property, which is perceptually desirable. In addition, the suboptimal solution of the pre- and post-filter can be obtained via Newton's algorithm. Computer simulation results show that this suboptimal solution does exhibit very good results in terms of improving the spectral matching and the output SNR. Furthermore, compared with the single optimal IS post-filter design, the jointly suboptimal pre- and post-filter is superior as expected.

CHAPTER V

REAL SPEECH SIMULATION

In this Chapter, we provide some experimental results on real speech signals. The experiments were conducted to compare the performance between the Wiener filter and the proposed optimal IS filter. The real speech data was obtained from the class ECEN 5753: Digital Speech Signal Processing, taught by Dr. K. Teague [Tea91]. Two English sentences are used:

1. "Thief who robs friend deserves jail," pronounced by a male speaker.
2. "Add the sum to the product of these three," pronounced by a female speaker.

All speech signals are sampled at the rate of 8K Hz. by a spectral analyzer interfaced with the workstation-Hypersignal DSP software. The acquired data is then converted into ASCII format to be compatible to the simulation programs written in the previous Chapters. Furthermore, for simulation purposes, the power of the speech sentence is normalized to be one.

The simulation is performed on a frame by frame basis. For simulation purposes, we restricted the frame size to be 128, which is commonly used in speech signal processing. The autocorrelation functions can be estimated from the time domain signal by any existing autocorrelation function estimator. We note that large

frame size will yield poor estimates of the autocorrelation function since speech is a time-varying model.

5.1 Simulation Results and Discussions

In this section, we perform computer simulations on both the single post filter and jointly optimal filter cases.

5.1.1 Single post-filter experiment

Two computer experiments are conducted according to the sentence one and two, respectively. The simulation scheme is that shown in Figure 3.1. Each experiment is carried out in two phases:

phase 1: vary the corrupting noise variance and fix the filter order.

phase 2: vary the filter order and fix the corrupting noise variance.

One different aspect of this simulation compared to the previous Chapters is the convergence criteria of the optimal IS filter. In the previous Chapters, we declared the optimal IS algorithm converged if

$$\xi_n - \xi_{n-1} \leq \epsilon , \quad (5.1)$$

where ξ_i is the IS distortion measure at the i^{th} iteration and ϵ is a small constant number.

However, in practice, it is difficult to estimate the IS distortion measure accurately from a small portion of speech signal. Furthermore, the discontinuity of the speech analysis frame causes the estimated power spectrum to fluctuate, resulting in

large number of erroneous IS distortion measurements. As a result, in this simulation, we propose a new convergence criteria as

$$\sum_{j=0}^k |\hat{R}_{xx}(j) - R_{xx}(j)| \leq \delta , \quad (5.2)$$

where k is an integer number, δ is a small constant number, and $|x|$ denotes the absolute value of x .

Experiment 1. In this experiment, sentence one is processed. The speech signal is corrupted by additive white Gaussian noise of known variance. With the sampling rate of 8K Hz., the total number of speech samples is 17920. Thus, there are 140 speech frames to be processed.

In the first phase, the filter order is fixed to be 5 and three different values of corrupting noise variances, σ^2 , are selected; 1, 2, and 3. These correspond to the input SNR of 1, 0.5, and 0.3333, respectively. The first 7680 samples of the original speech is depicted in Figure 5.1. The first 7680 samples of the output of single Wiener filter and the optimal IS filter for the case of $\sigma^2 = 2$ are shown in Figure 5.2 and 5.3, respectively. Comparing Figure 5.2 and 5.3, notice the loss of signal energy in the Wiener filter output (for example from the sample 500 to sample 3000), compared to the output of the optimal IS filter. The IS filter output is more closely matches the original speech. This is because the optimal IS filter has the autocorrelation function matching property while the Wiener filter does not. To emphasize the improvement of the optimal IS filter over the Wiener filter and for comparison purpose, we define the autocorrelation function error of the i^{th} frame, E_i , as

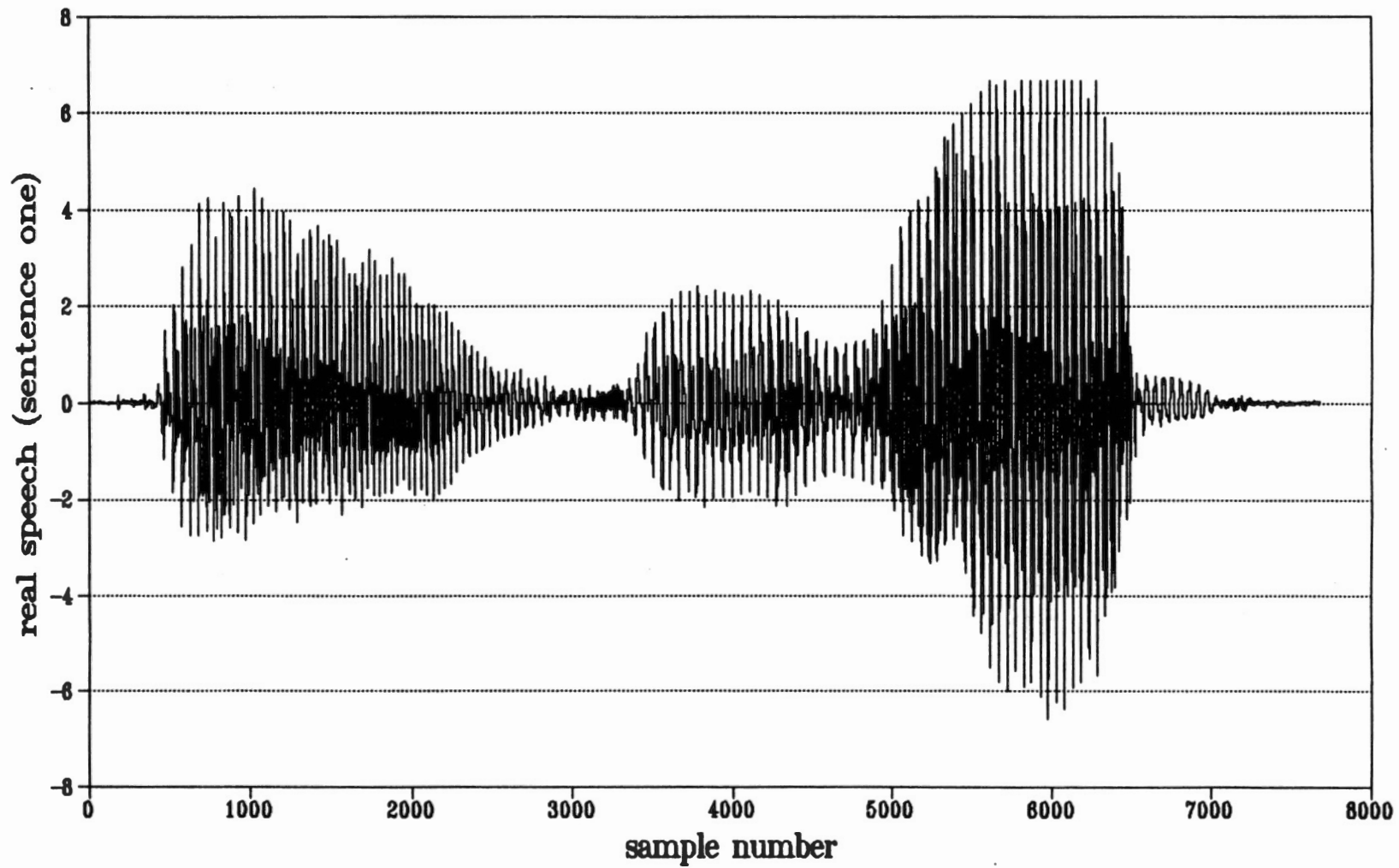


Figure 5.1 Plot of the First 7680 Samples of Sentence One.

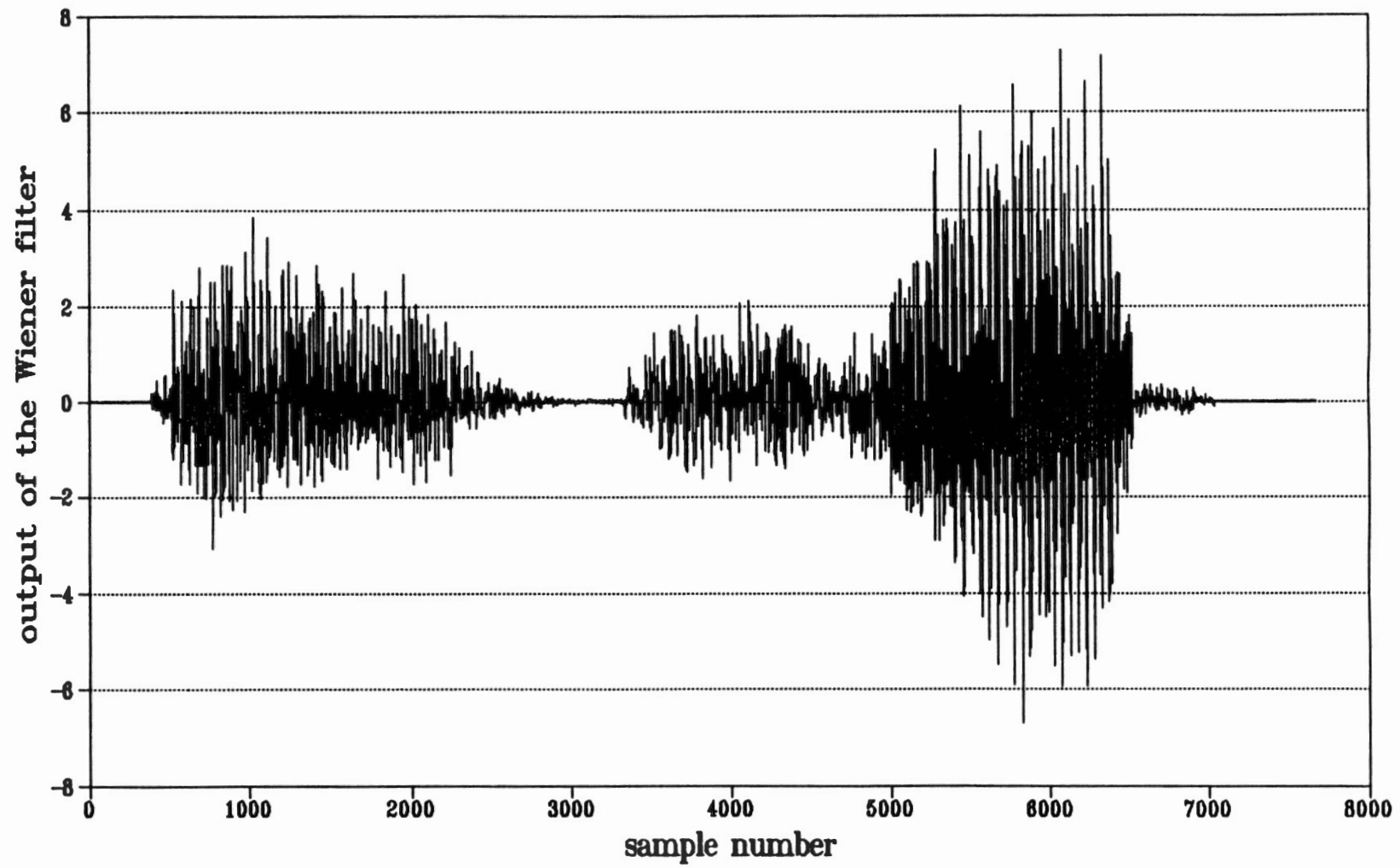


Figure 5.2 Plot of the First 7680 Samples of the Wiener Filter Output.

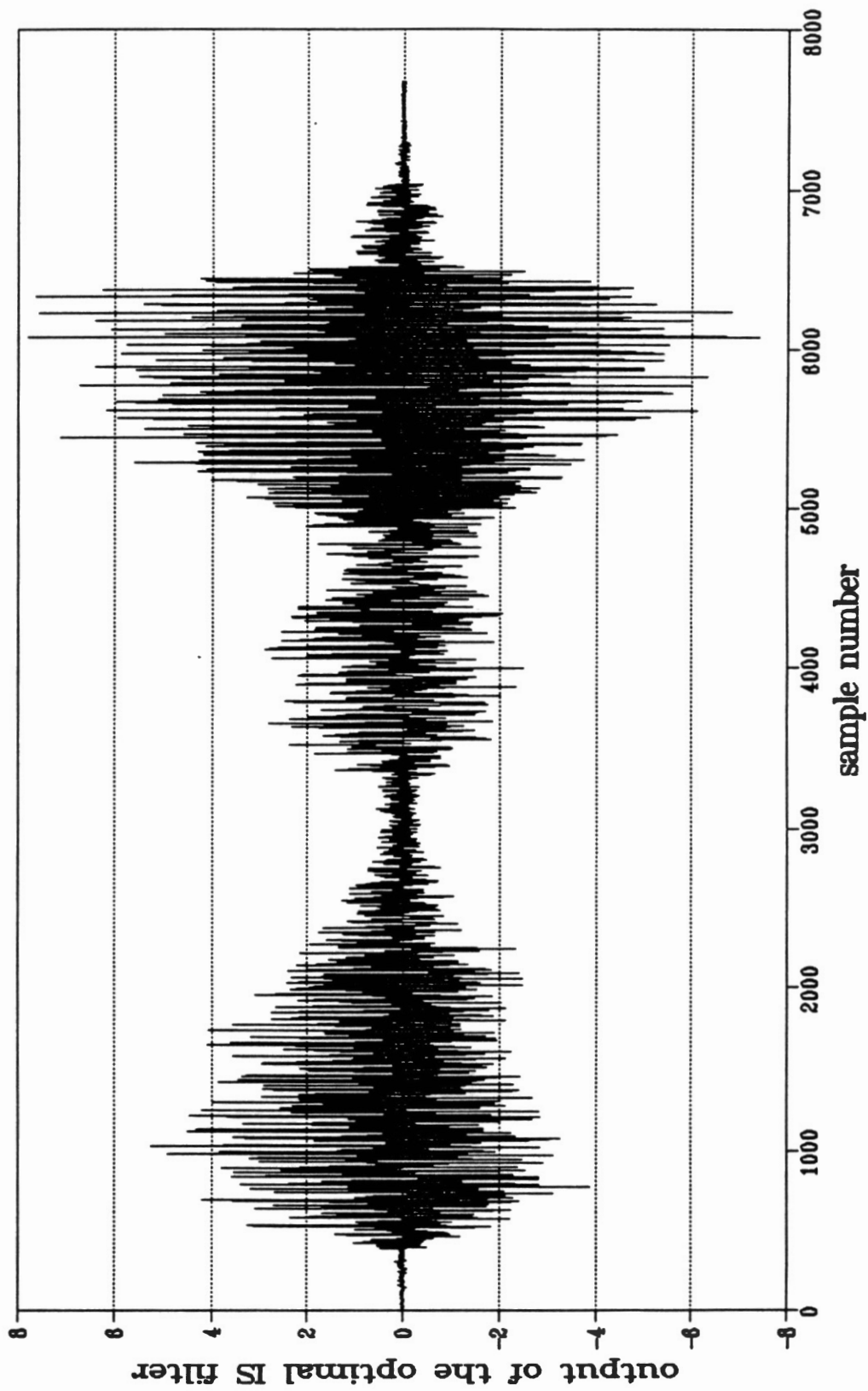


Figure 5.3 Plot of the First 7680 Samples of the Optimal IS Filter Output.

$$E_i = \sum_{j=0}^{10} |\hat{R}_{xx}(j) - R_{xx}(j)| \quad (5.3)$$

Note that equation (5.3) implies k in equation (5.2) is equal to 10. Large value of E_i implies large error in autocorrelation function matching in the i^{th} frame, and vice versa. In Figure 5.4, 5.5, 5.6, we compare the autocorrelation function error of the Wiener filter and that of the optimal IS filter as a function of the frame number for the cases where $\sigma^2 = 1, 2, \text{ and } 3$, respectively. From these Figures, we can see that the autocorrelation function error of the optimal IS filter are always less than the autocorrelation function error of the Wiener filter for all three values of the corrupting noise variances. This implies that the autocorrelation functions of the output of the optimal IS filter are closer to the autocorrelation functions of the real speech signal than the autocorrelation functions of the Wiener filter output. Figures 5.7, 5.8, and 5.9 show the comparison in terms of the IS distortion measure versus the frame number for all three corrupting noise variance. We can see that the IS distortion measure of the optimal IS filter tends to be less than that of the Wiener filter. Thus, the optimal IS filter outperforms the Wiener filter in terms of both autocorrelation function error and the IS distortion measure.

In Figure 5.10, we compare the autocorrelation error of the optimal IS filter under three different values of corrupting noise variances. We can see from Figure 5.10 that, for any specific speech frame, the autocorrelation function error increases as the noise variance increases. This is expected since as discussed in the previous Chapters, the ability to match the autocorrelation function of the optimal IS filter

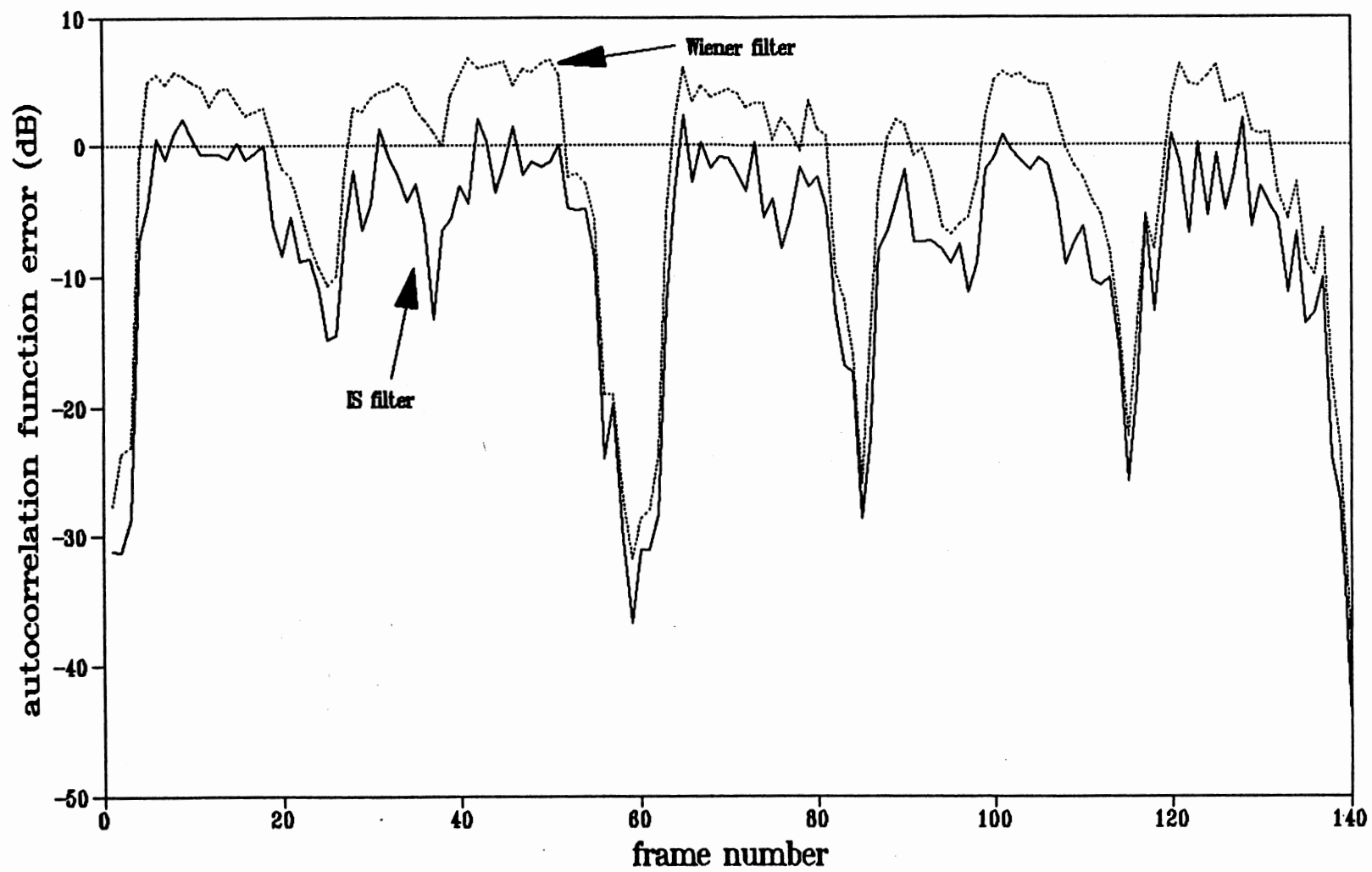


Figure 5.4 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 1$).

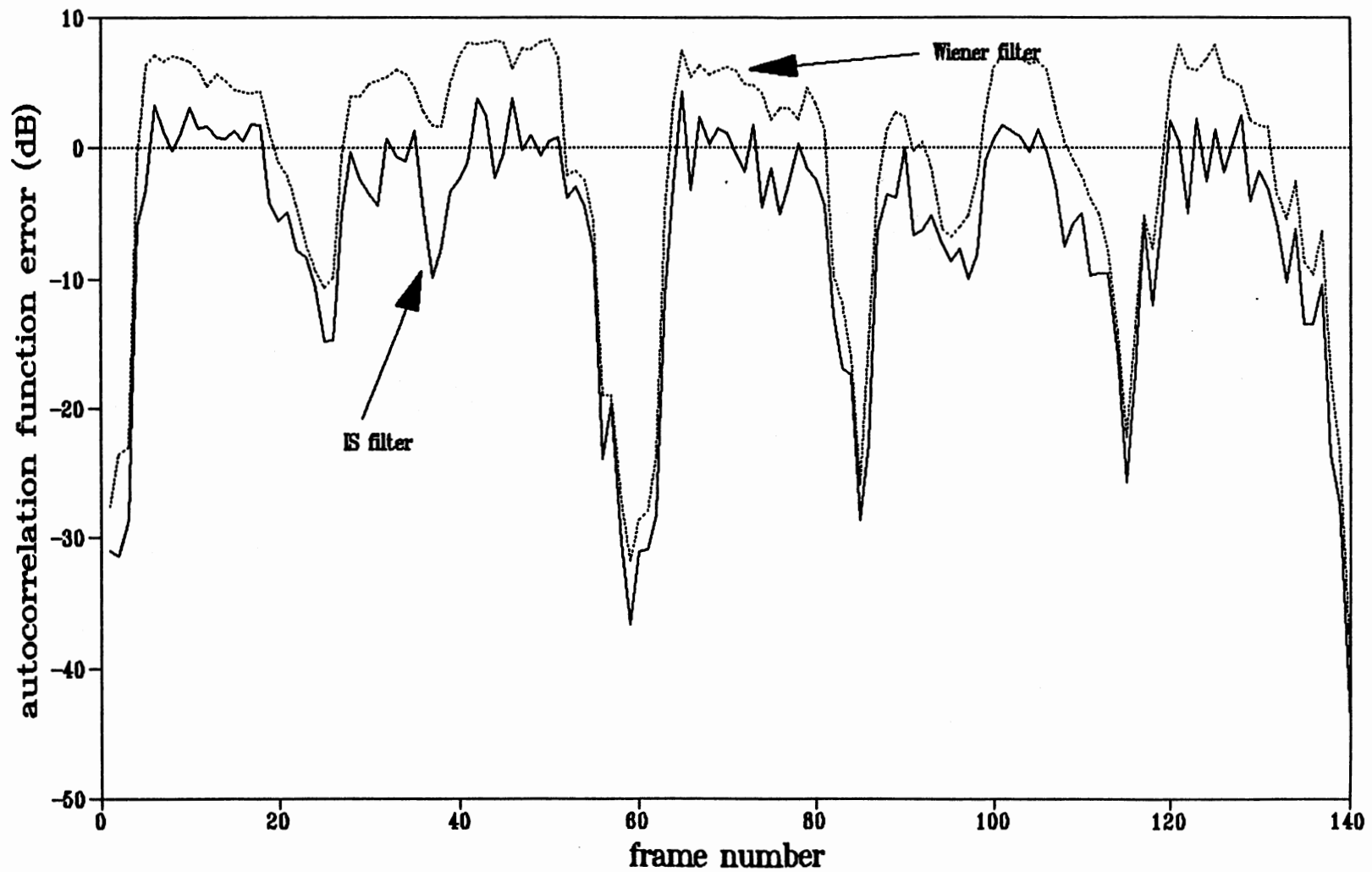


Figure 5.5 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 2$).

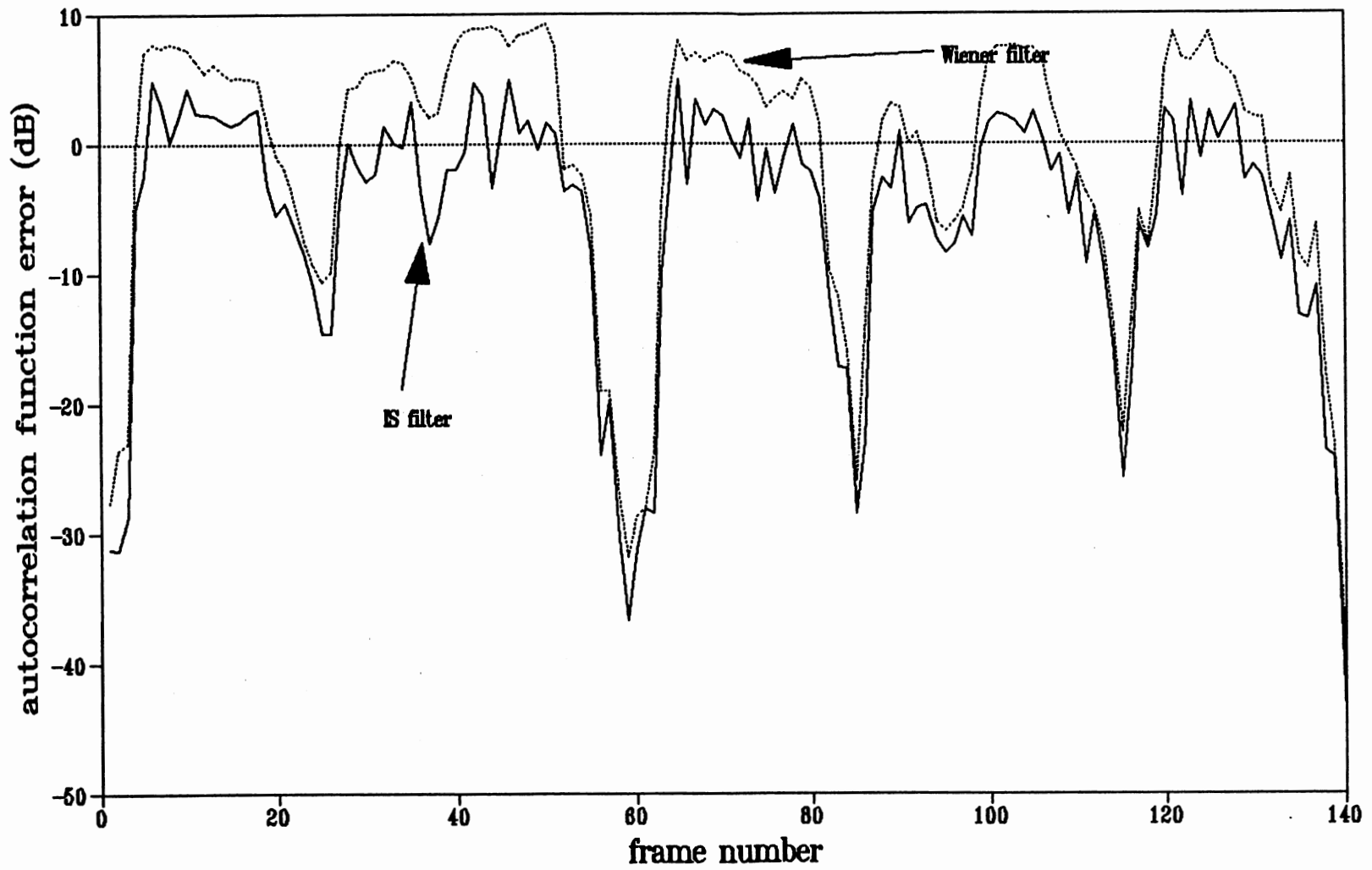


Figure 5.6 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2 = 3$).

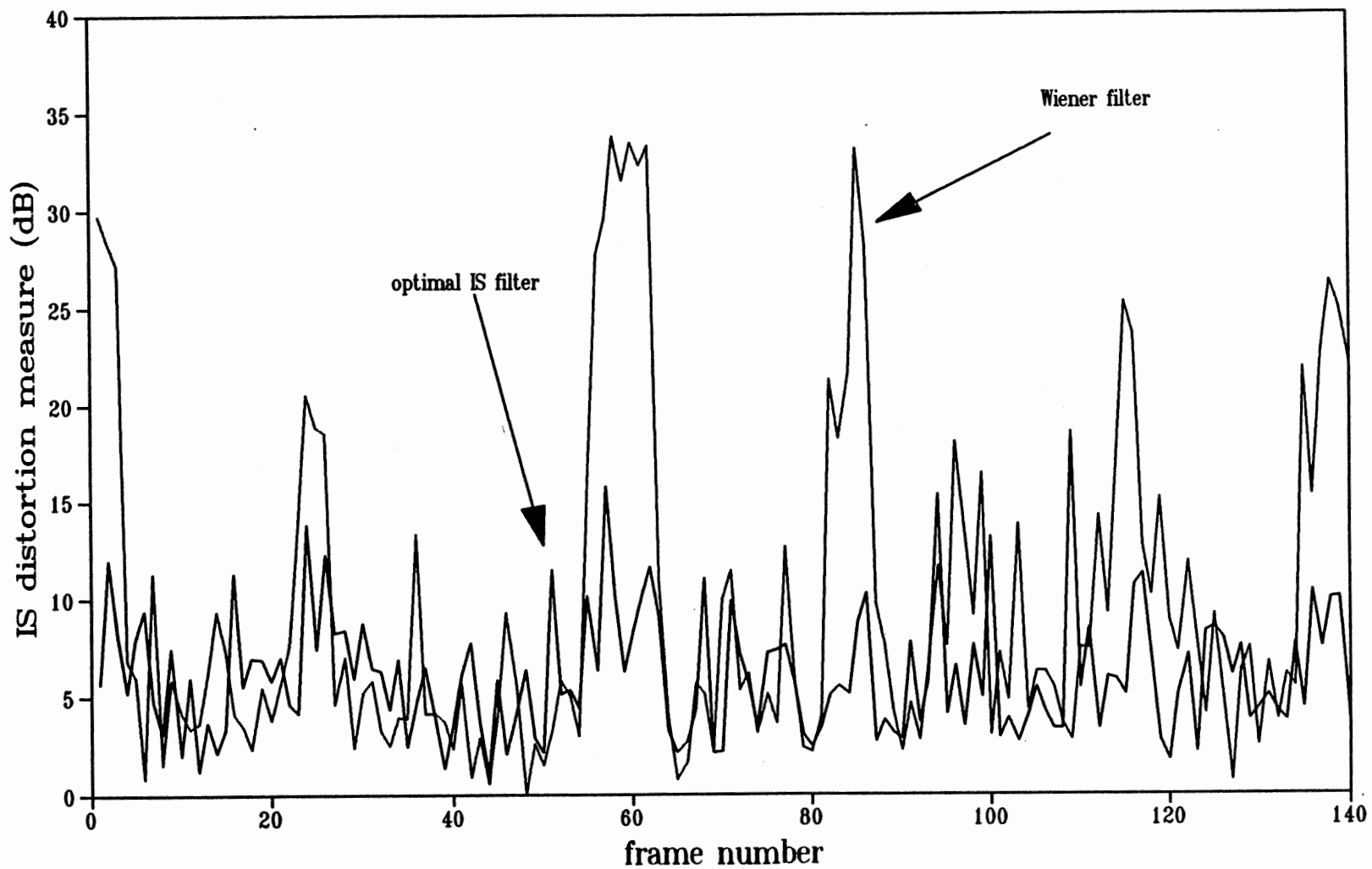


Figure 5.7 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 1$).

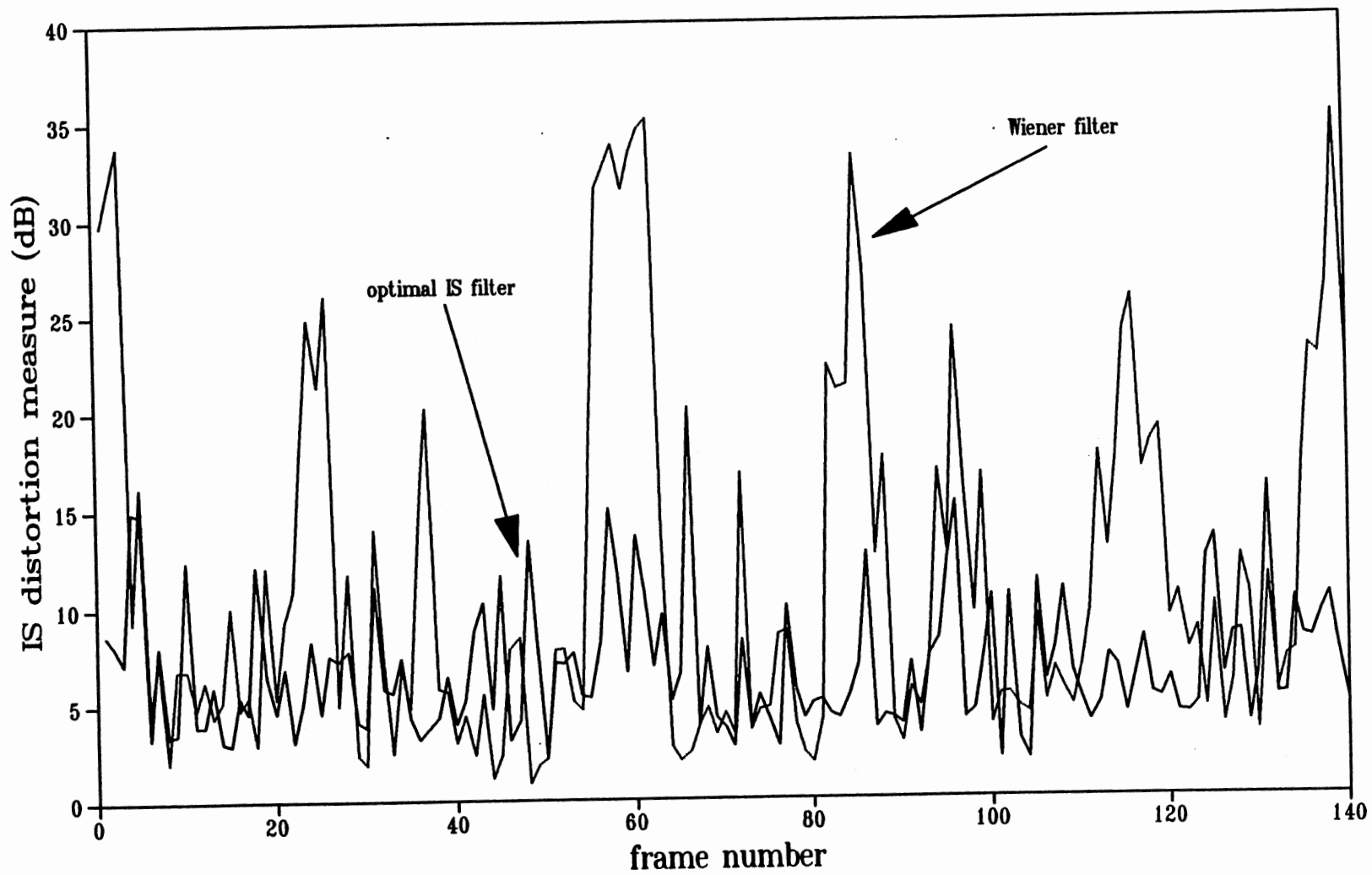


Figure 5.8 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 2$).

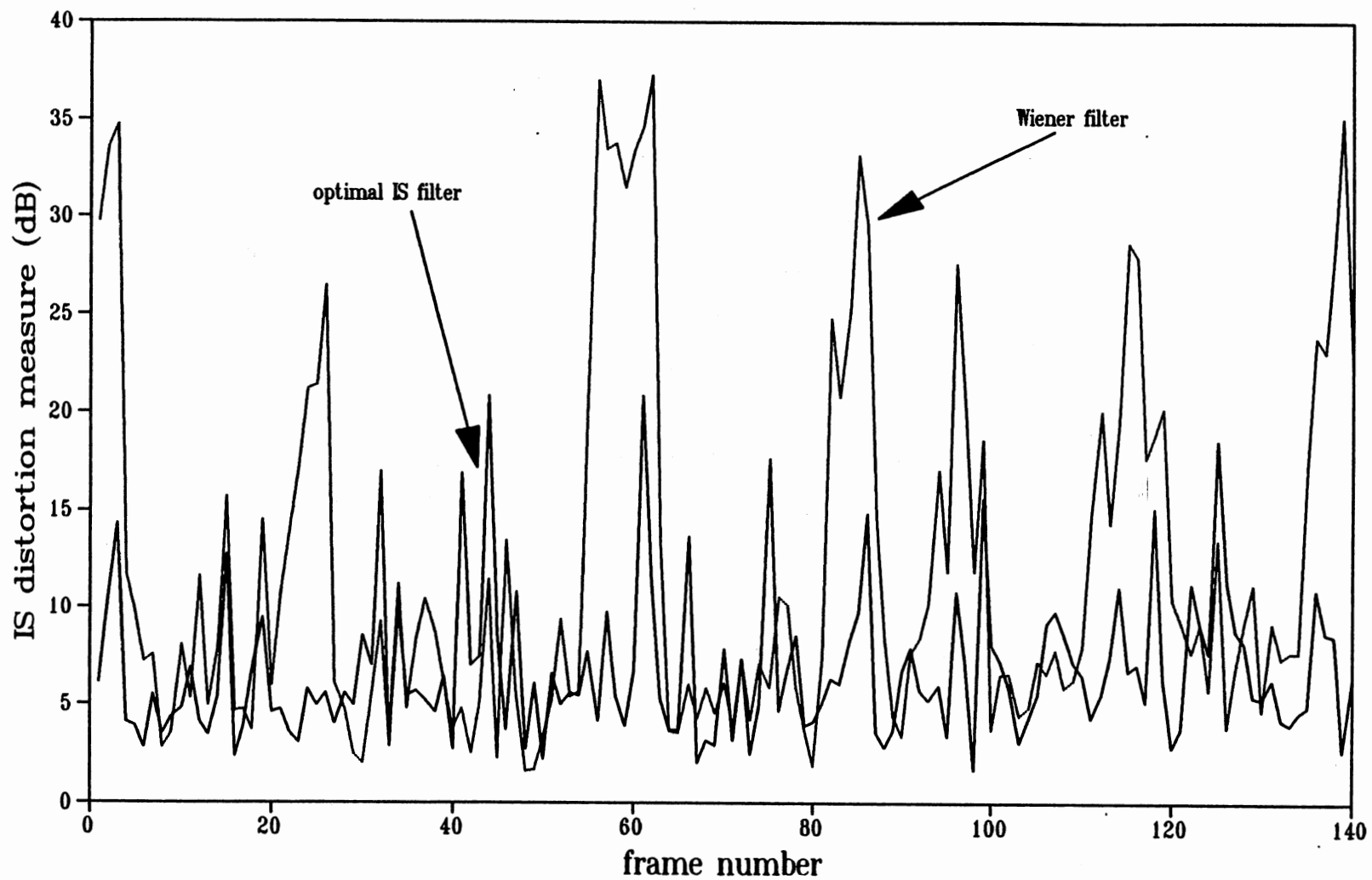


Figure 5.9 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 3$).

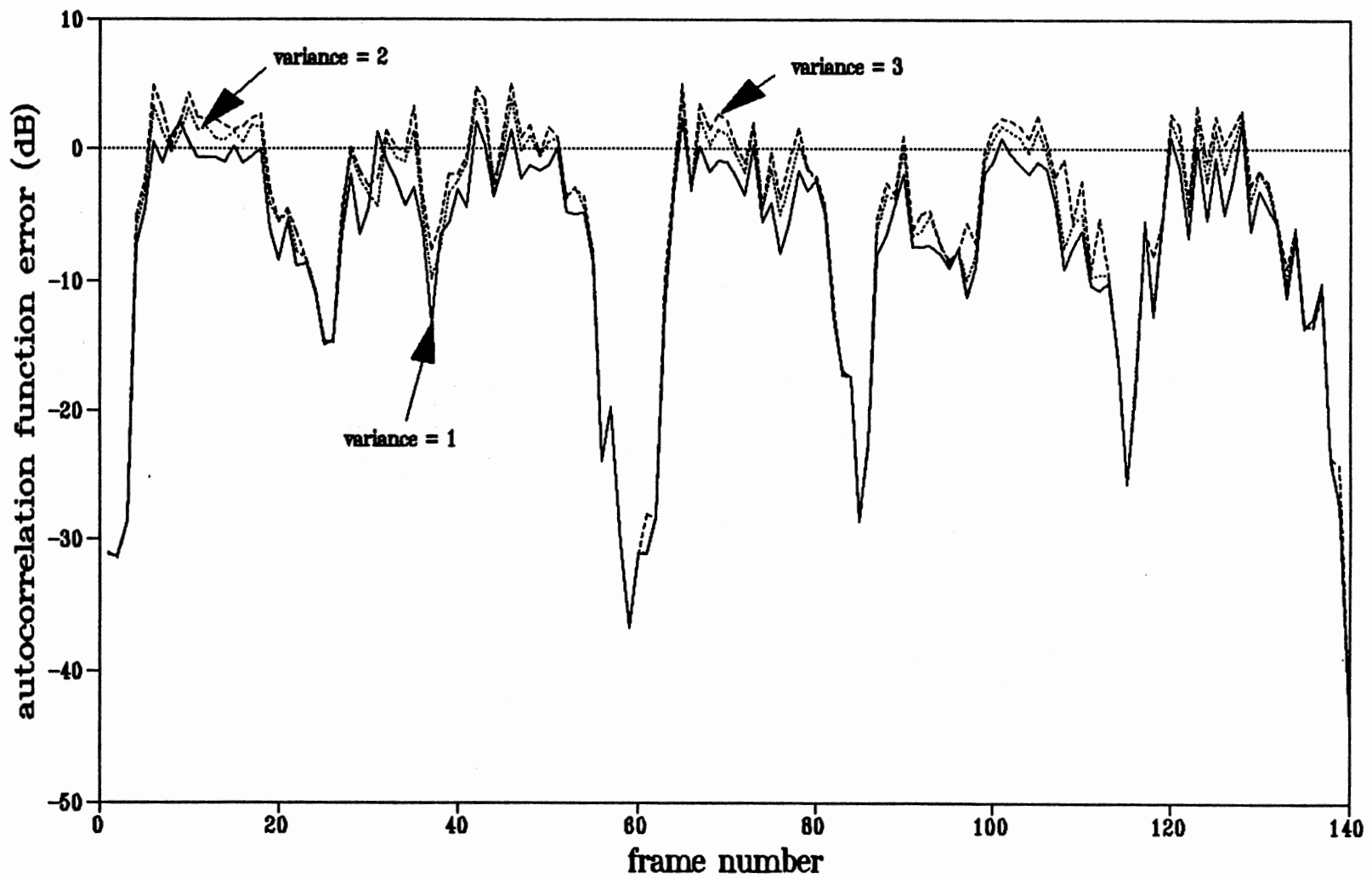


Figure 5.10 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different Noise Variances.

decreases as the noise variance increases. Furthermore, note that the output of the optimal IS filter is more peaky than the output of the Wiener filter. This implies the optimal IS filter has higher output power than the Wiener filter which coincides with the discussion in the previous Chapters.

In the second phase, the corrupting noise variance is fixed to be one and the FIR filter order is varied between 2, 5, and 10. To the extreme case, we plot the first 7680 samples of the output of the optimal IS filter for the cases of filter orders are equal to 2 and 10 in Figure 5.11 and 5.12, respectively. From Figure 5.11 and 5.12, there is no significant different in terms of signal resemblance to the original speech (Figure 5.2). However, Figure 5.13 compares the autocorrelation function error as a function of the frame number of all three FIR filter orders. From Figure 5.13, we can see that, for any specific frame, the autocorrelation function error tends to decrease as the filter order increases, which implies improvement in autocorrelation function matching. This result is simply because as the filter order increases, more autocorrelation function lags can be matched to the autocorrelation function of the original speech.

Experiment #2. In this experiment, sentence two is also corrupted by additive white Gaussian noise of known variance. The total number of samples is 21888, resulting in 171 analysis frames to be processed. Note that a female speaker tends to have a higher pitched speech signal and narrower bandwidth. As in the previous experiment, in phase one the filter order is fixed to be 5 and the corrupting noise variance is varied between 1, 2, and 3, respectively. The first 7680 samples of

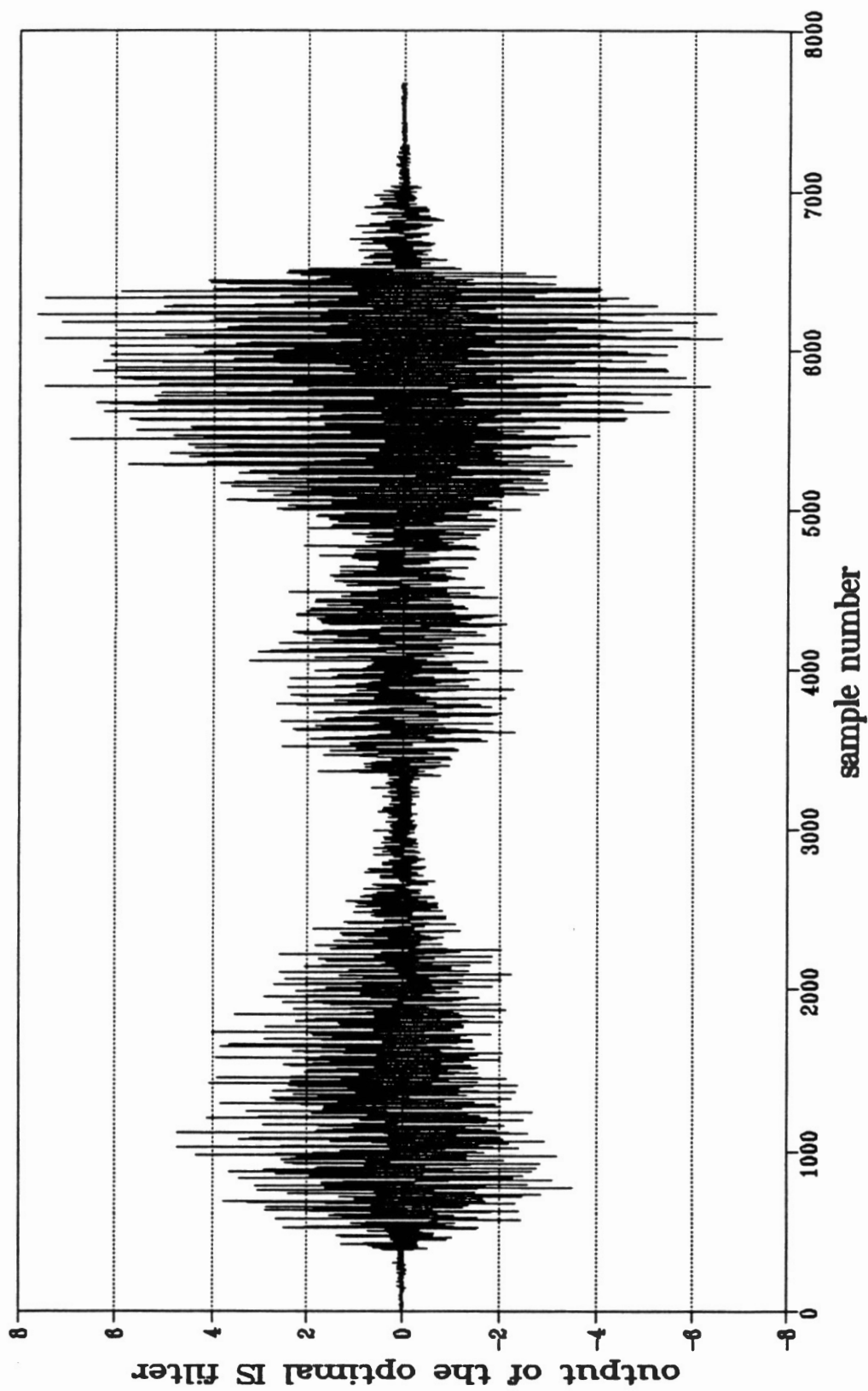


Figure 5.11 Plot of the First 7680 Samples of the Optimal IS Filter Output (Order = 2).

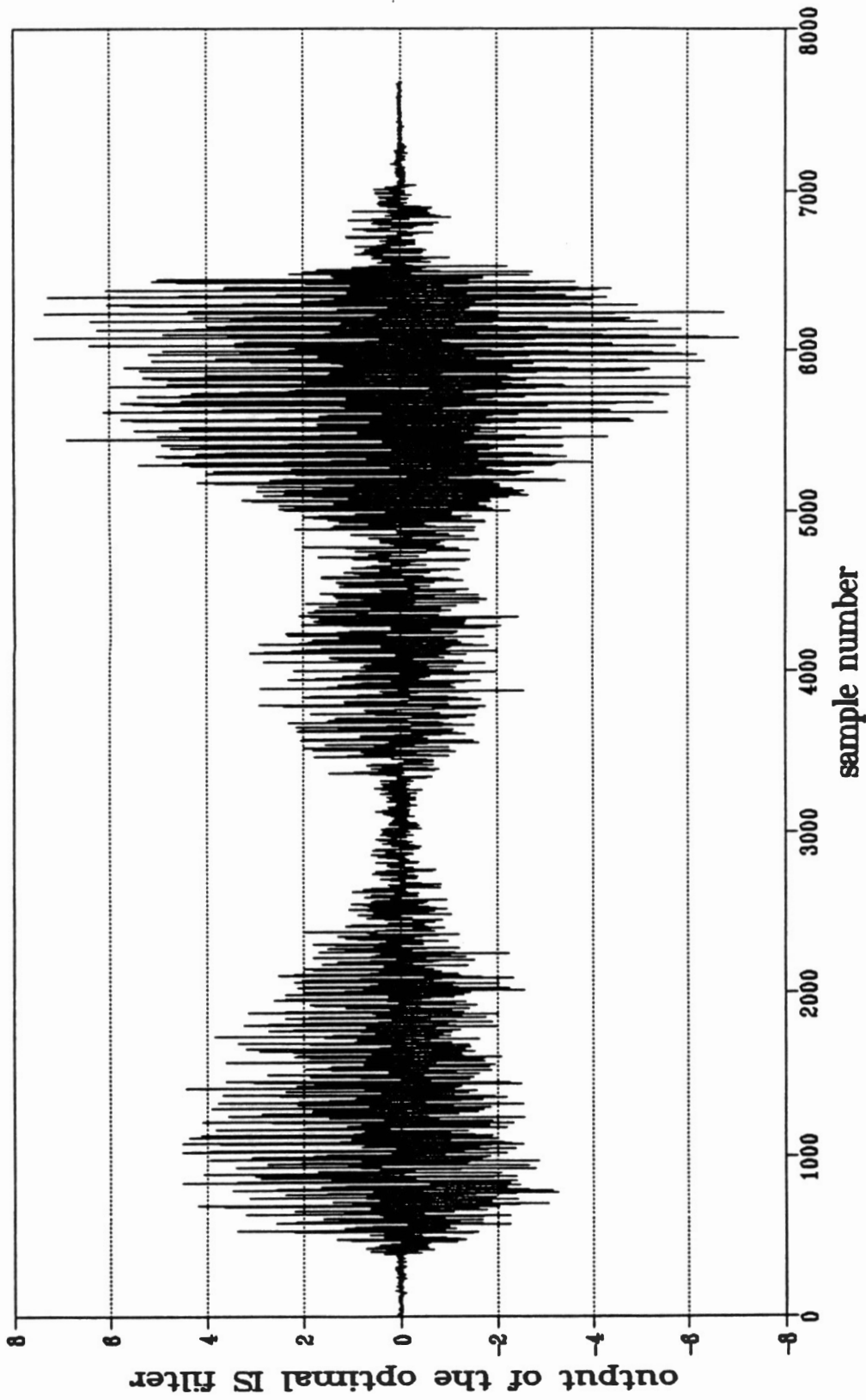


Figure 5.12 Plot of the First 7680 Samples of the Optimal IS Filter Output (Order = 10).

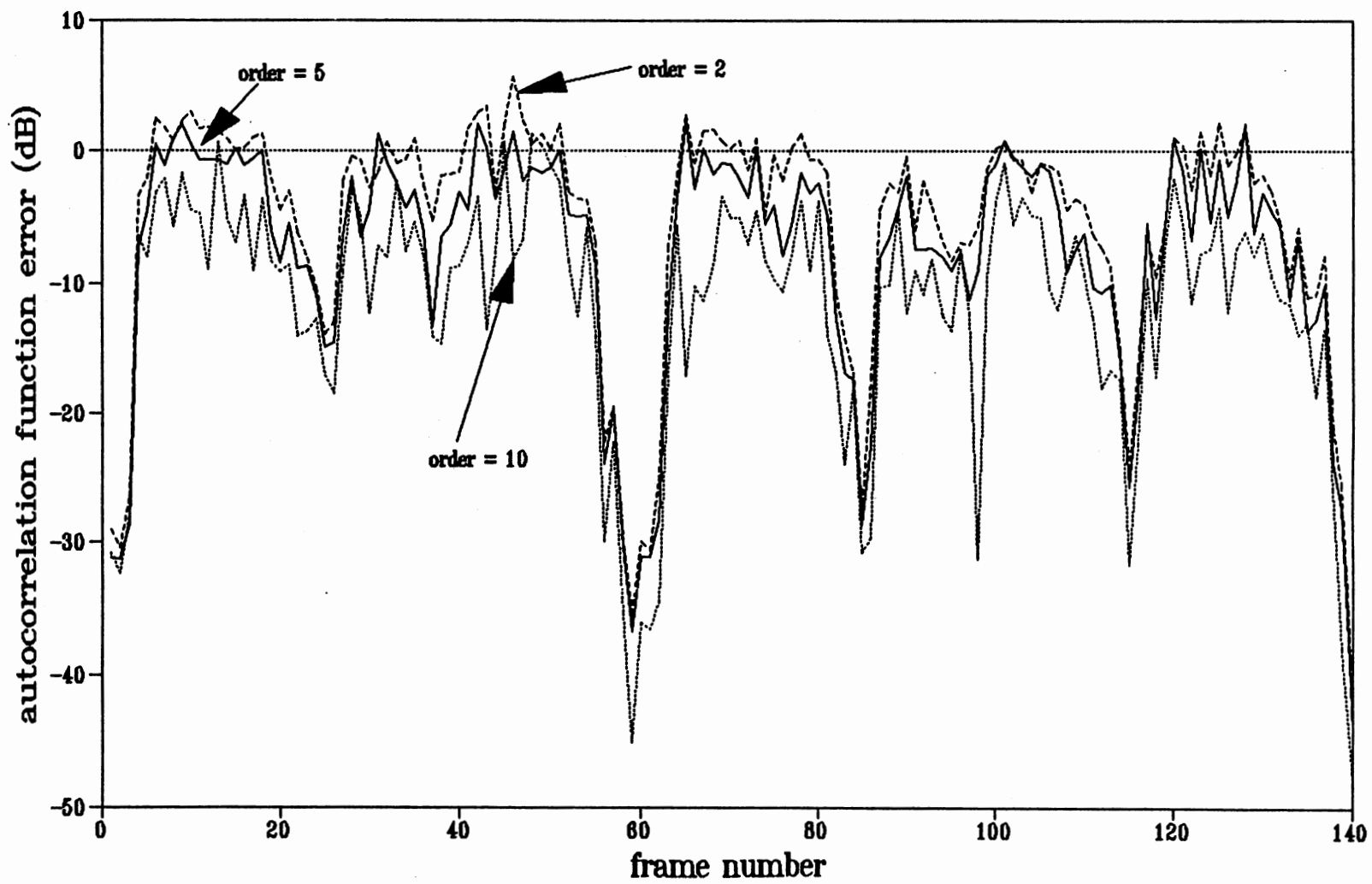


Figure 5.13 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different FIR Filter Orders.

the original speech is shown in Figure 5.14. The first 7680 samples of the output of the Wiener filter and the optimal IS filter in the case where $\sigma^2 = 2$ are depicted in Figure 5.15 and 5.16, respectively. Comparing Figure 5.15 with 5.16, again notice the larger loss of signal energy in the Wiener filter output compared to the optimal IS filter output. In Figure 5.17, 5.18, and 5.19, we compare the autocorrelation function error of the optimal IS filter and that of the Wiener filter as a function of the frame number for all three values of the corrupting noise variances, respectively. The autocorrelation function error of the optimal IS filter is always less than the autocorrelation function error of the Wiener filter for all three variances. Figure 5.20, 5.21, and 5.22 compare the IS distortion measure of the Wiener filter with that of the optimal IS filter for all three noise variance cases. We can see that the IS distortion measure of the optimal IS filter tends to be less than that of the Wiener filter. Thus, the optimal IS filter outperforms the Wiener filter in terms of the autocorrelation function matching and minimizing the IS distortion measure.

In Figure 5.23, we compare the autocorrelation function error of the optimal IS filter as a function of the frame number for all three values of the corrupting noise variances. From Figure 5.23, we can see that, for any specific frame, the autocorrelation function error increases as the corrupting noise variance increases. This implies that as the noise variance increases, the autocorrelation function of the optimal IS filter output tends to less resemble the autocorrelation function of the original speech.

In the second phase, the corrupting noise variance is fixed to be one and the

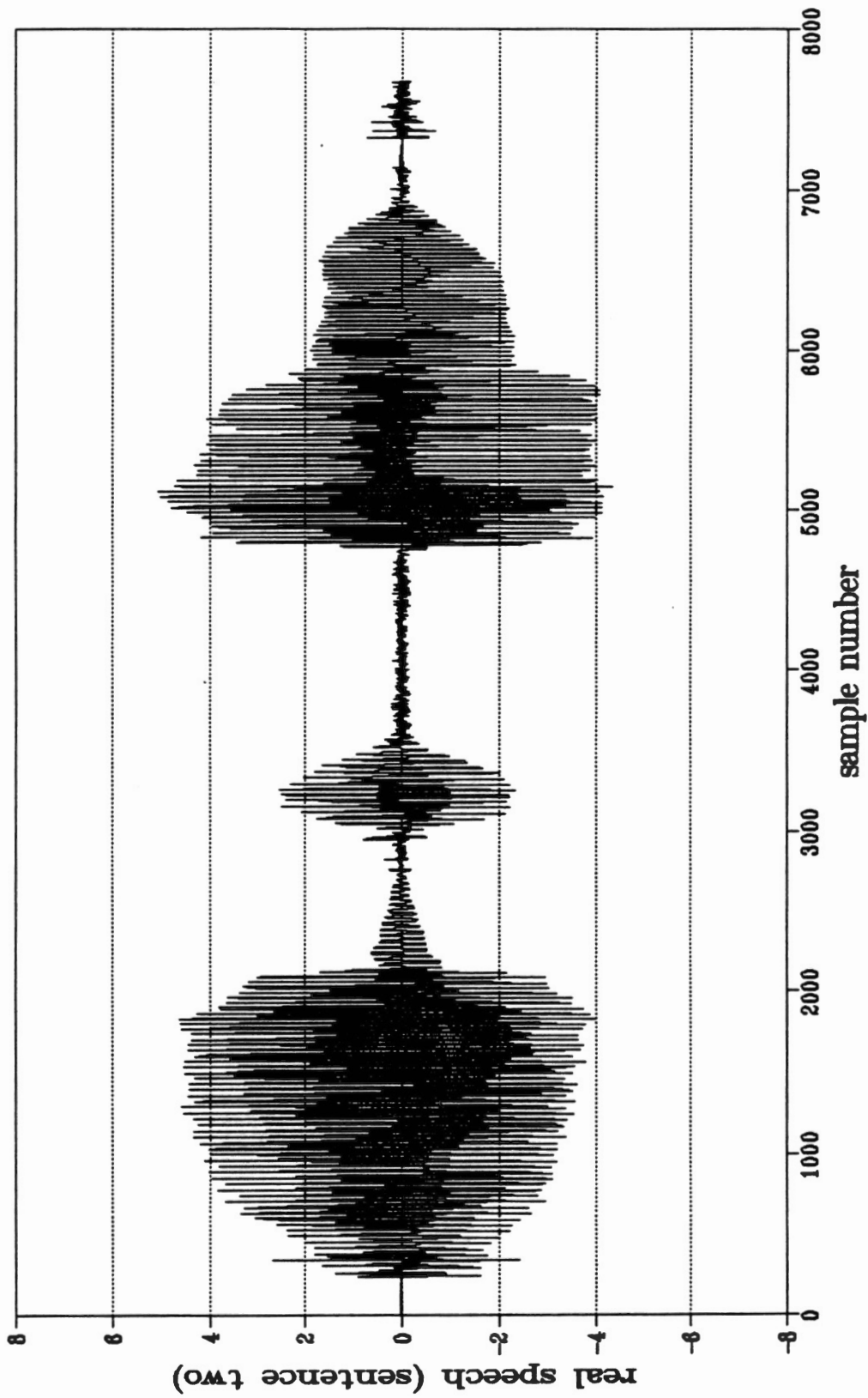


Figure 5.14 Plot of the First 7680 Samples of Sentence Two.

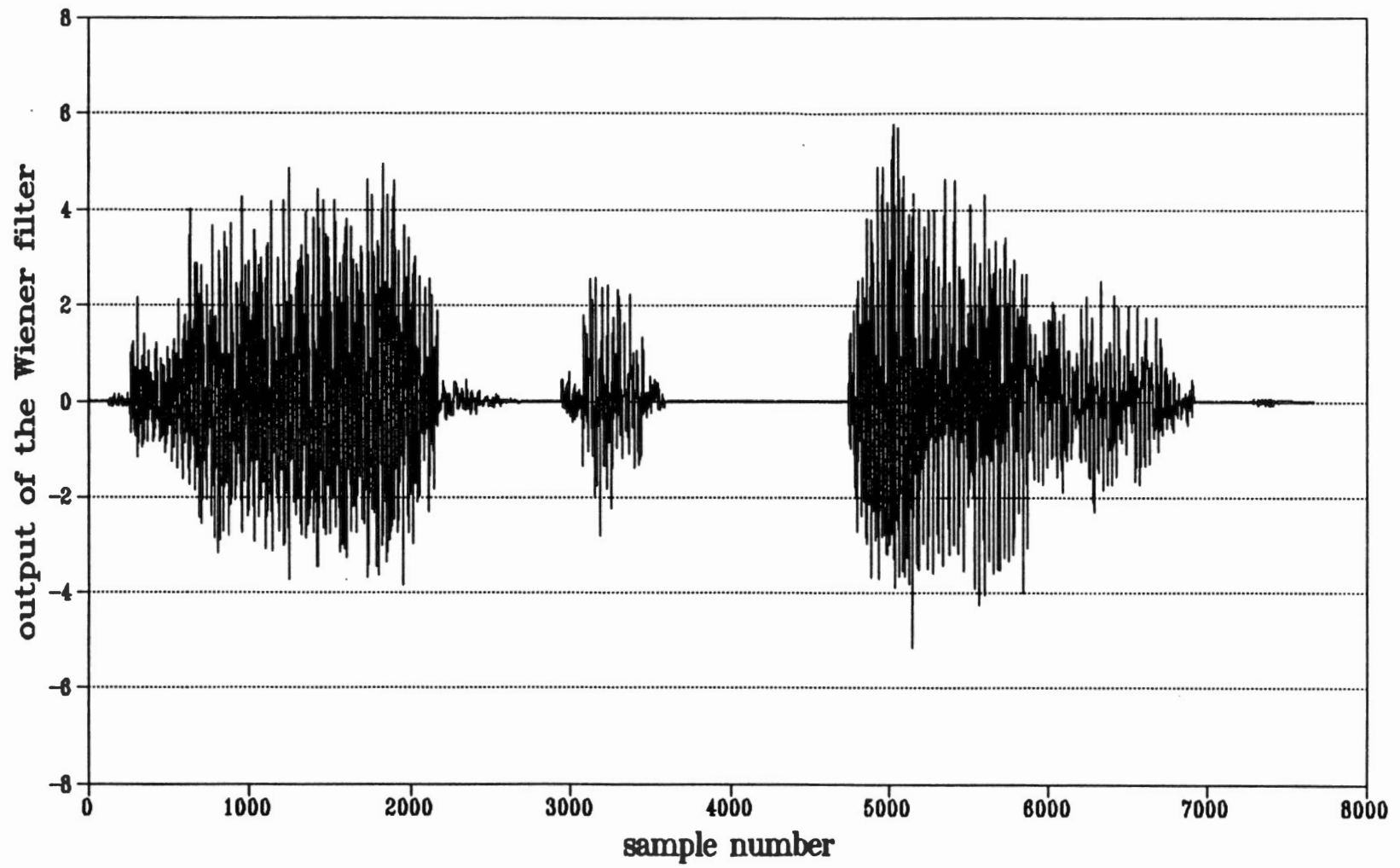


Figure 5.15 Plot of the First 7680 Samples of the Wiener Filter Output.

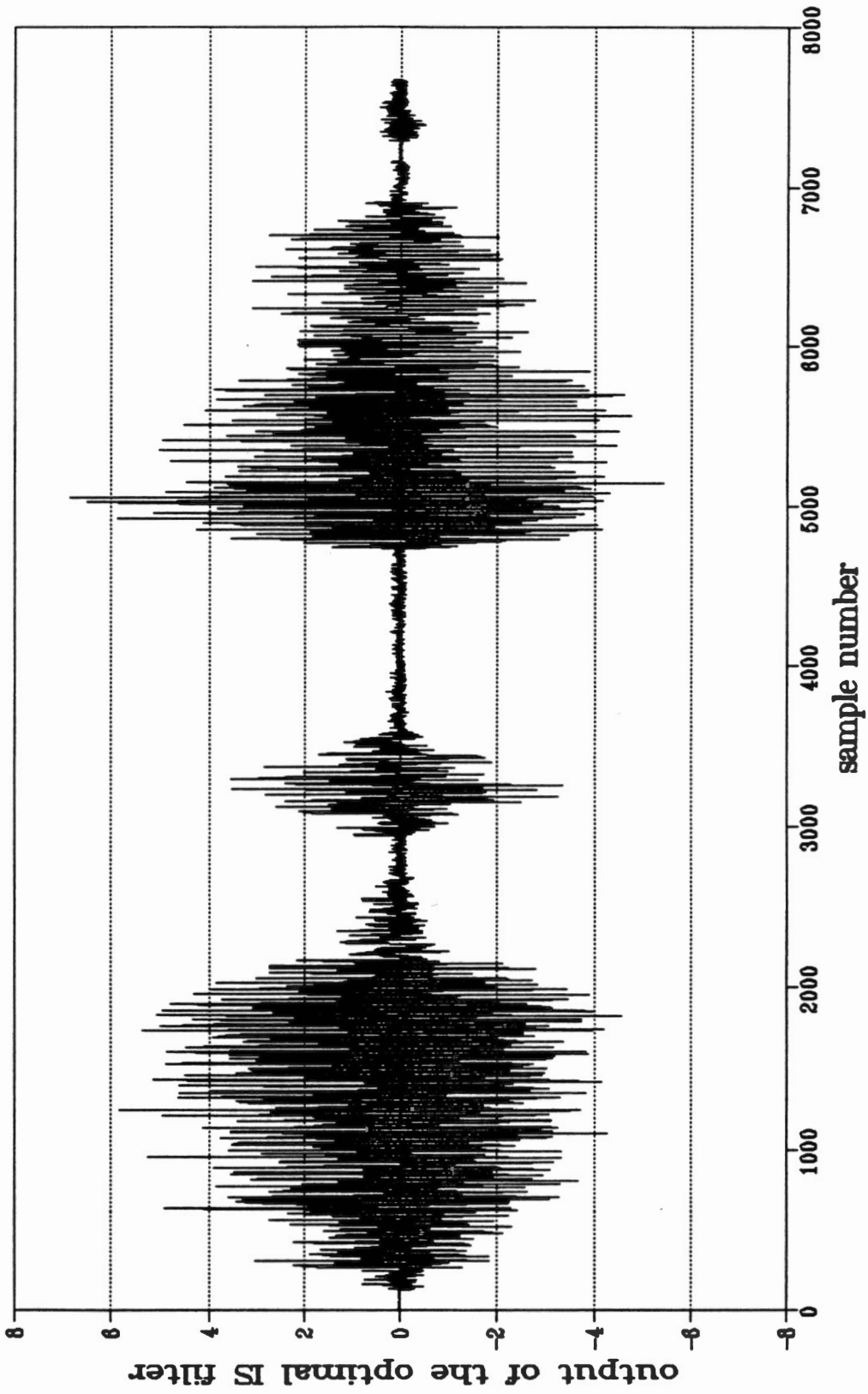


Figure 5.16 Plot of the First 7680 Samples of the Optimal IS filter Output.

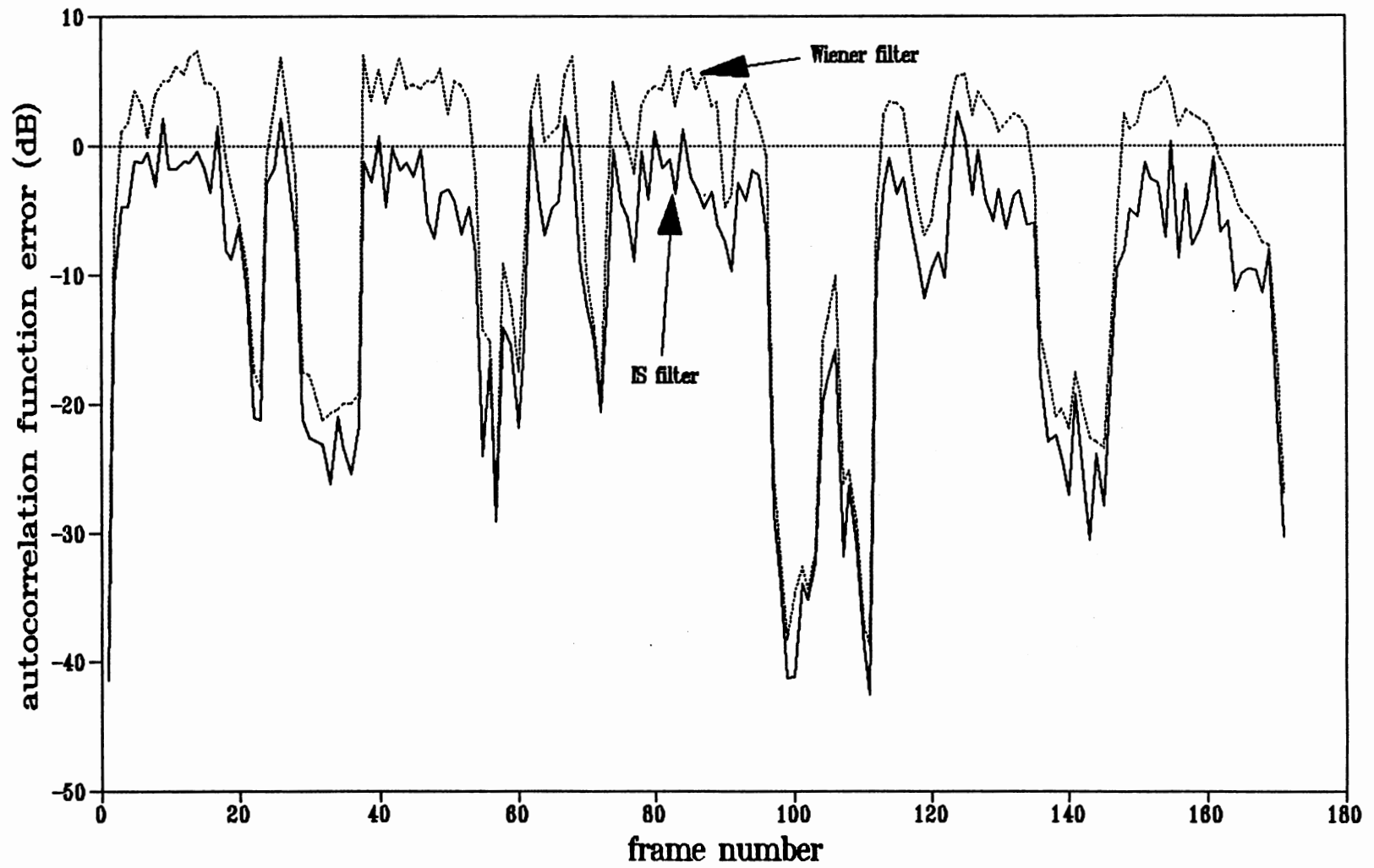


Figure 5.17 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2=1$).

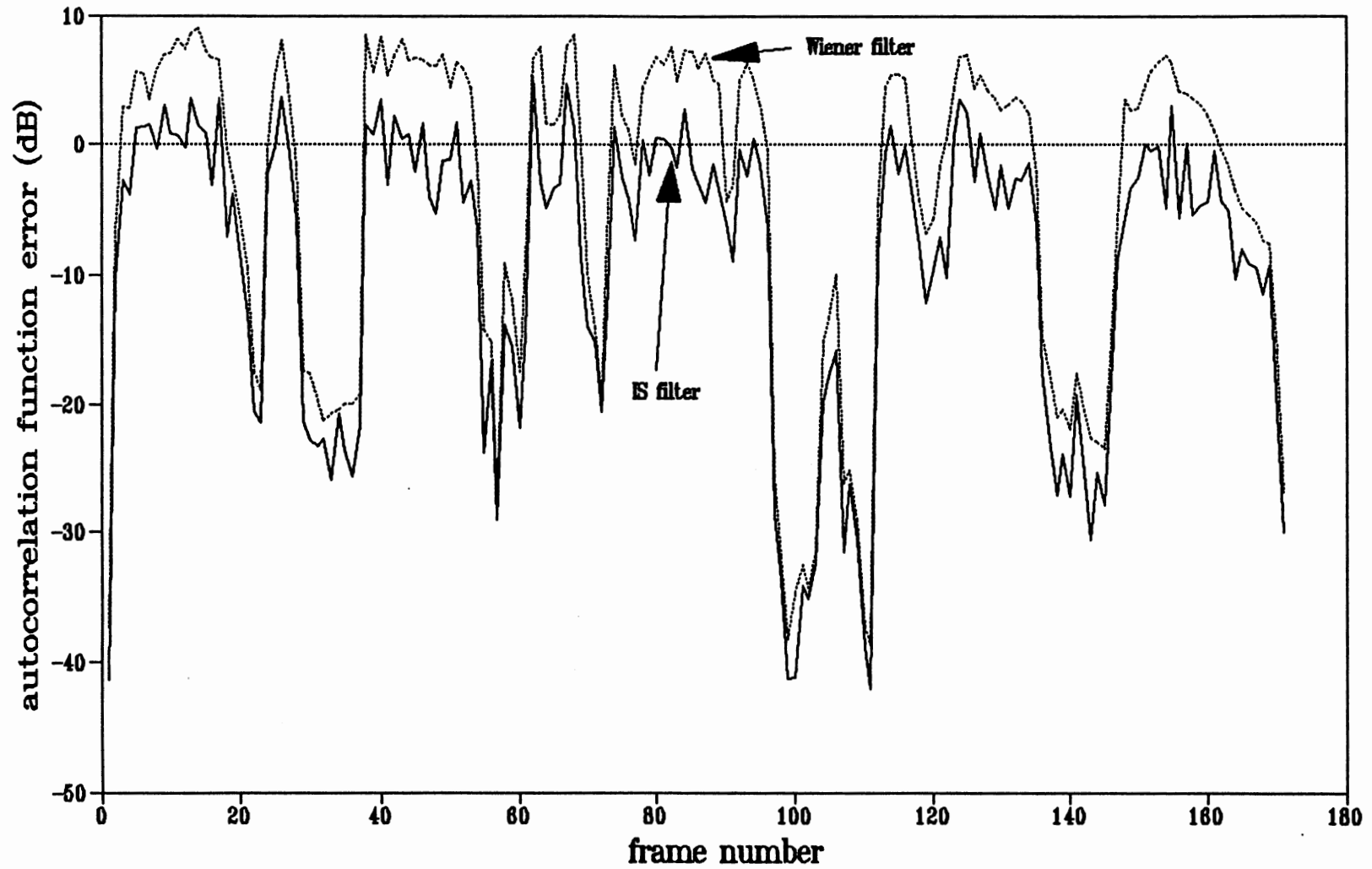


Figure 5.18 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2= 2$).

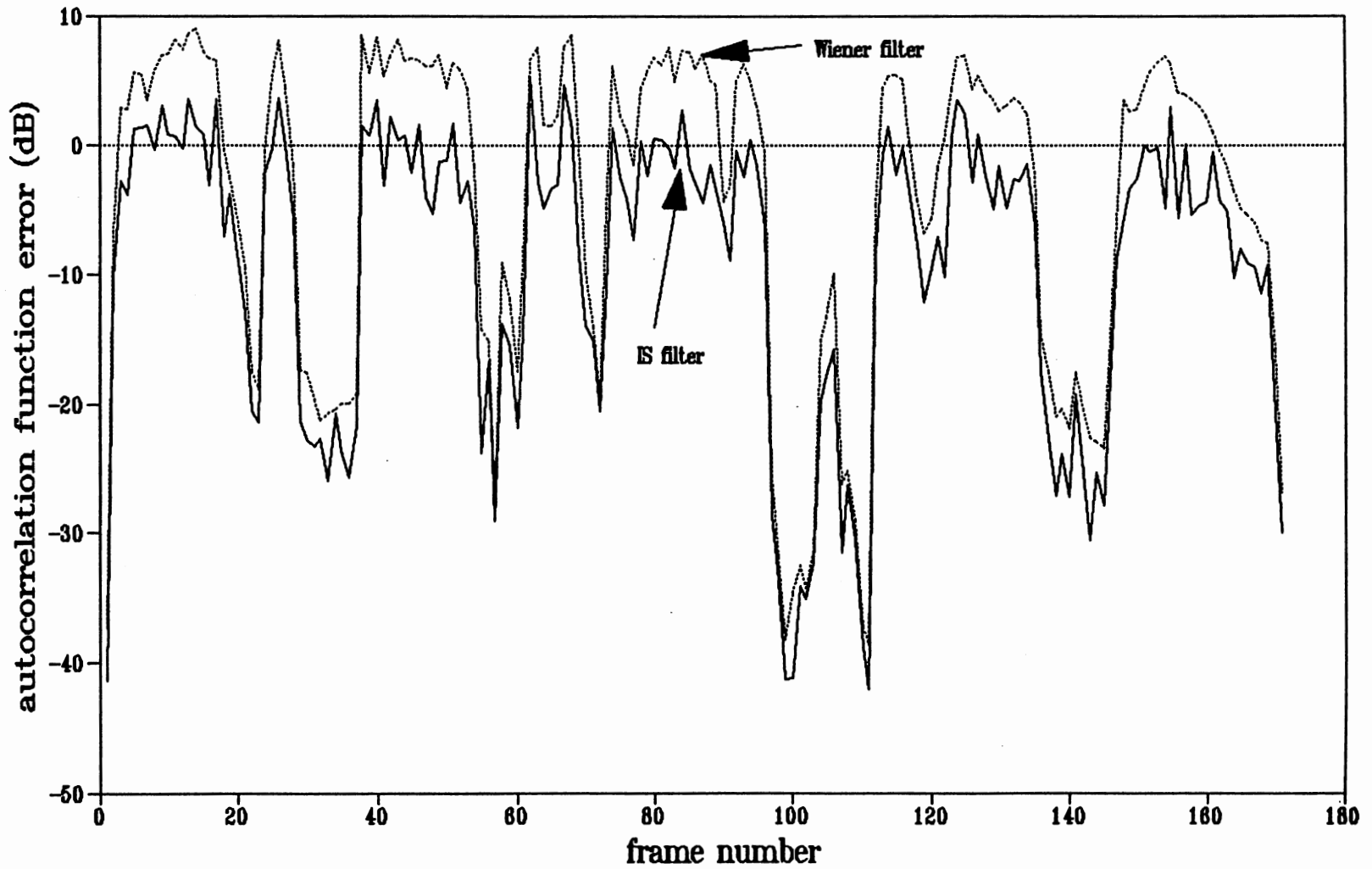


Figure 5.19 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the Autocorrelation Function Error ($\sigma^2=3$).

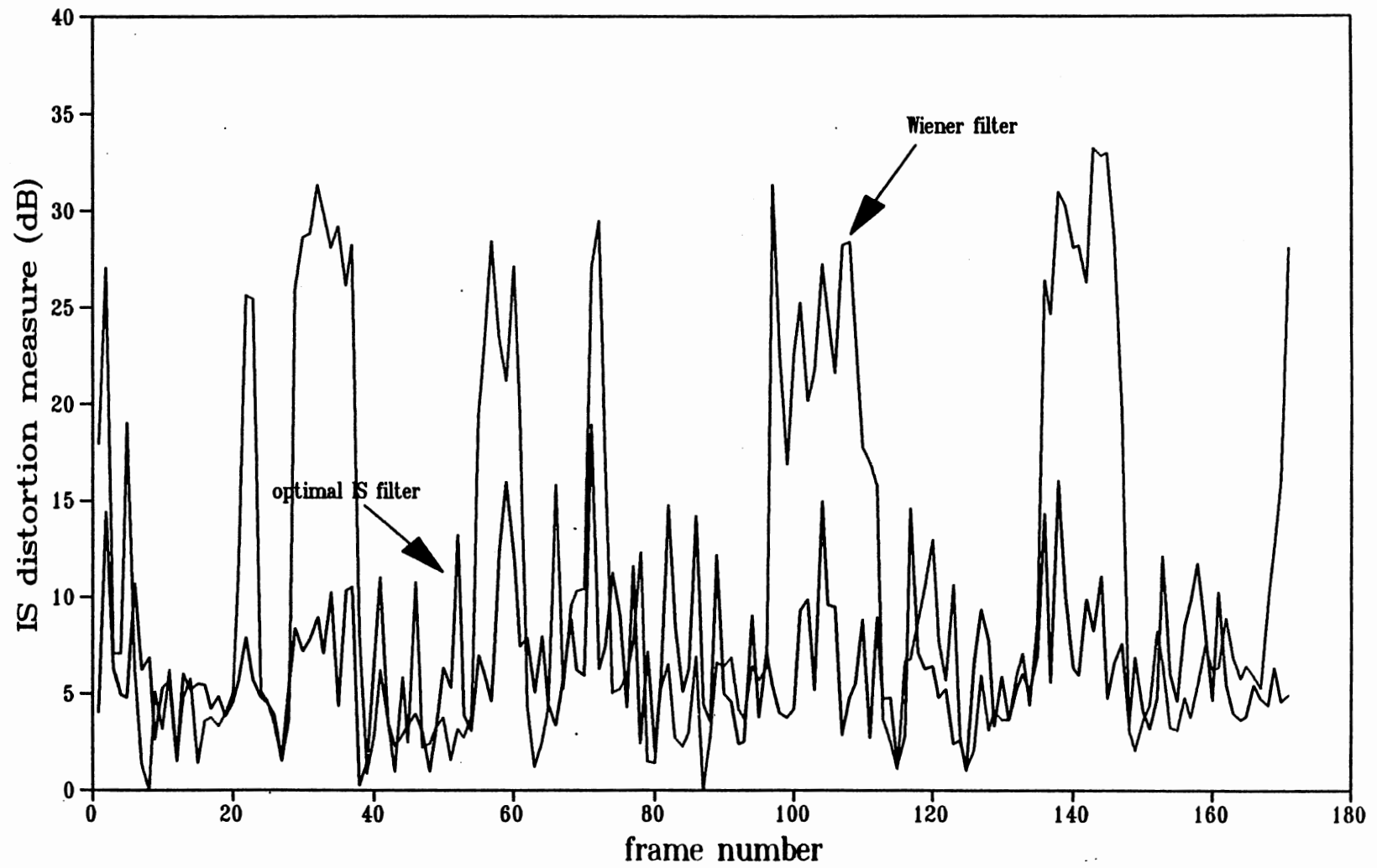


Figure 5.20 Comparison of the Wiener filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2=1$).

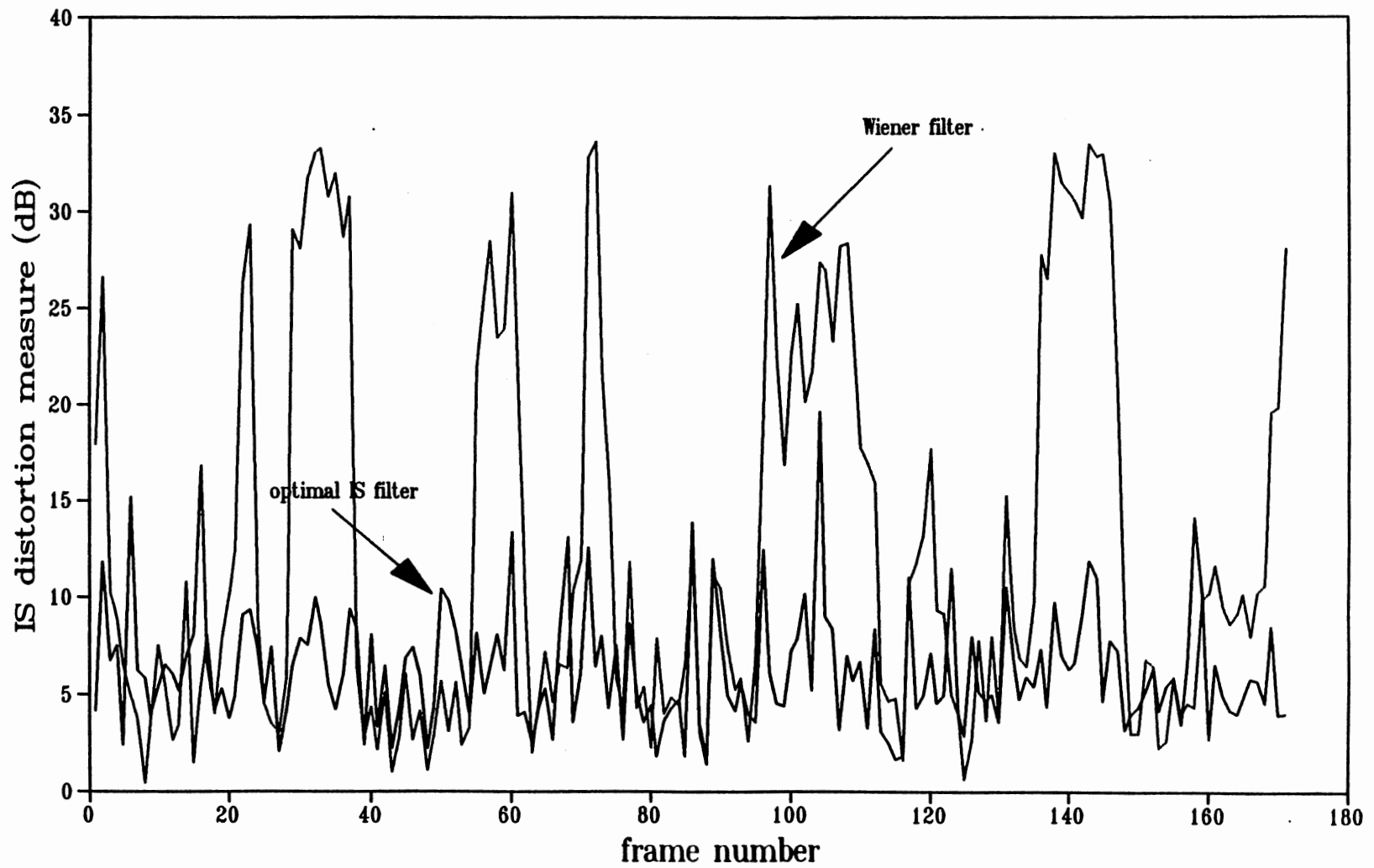


Figure 5.21 Comparison of the Wiener filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2=2$).

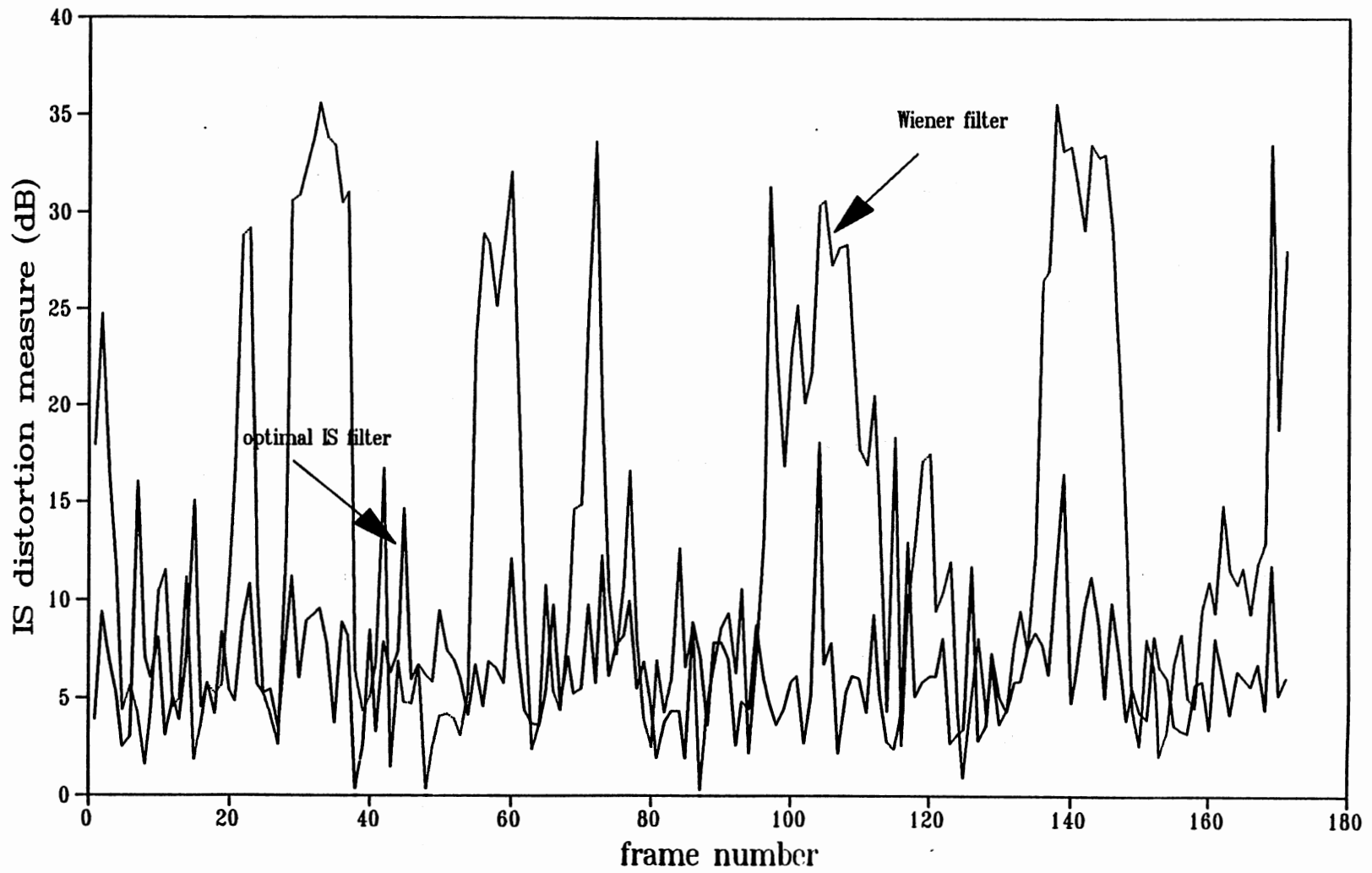


Figure 5.22 Comparison of the Wiener filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2=3$).

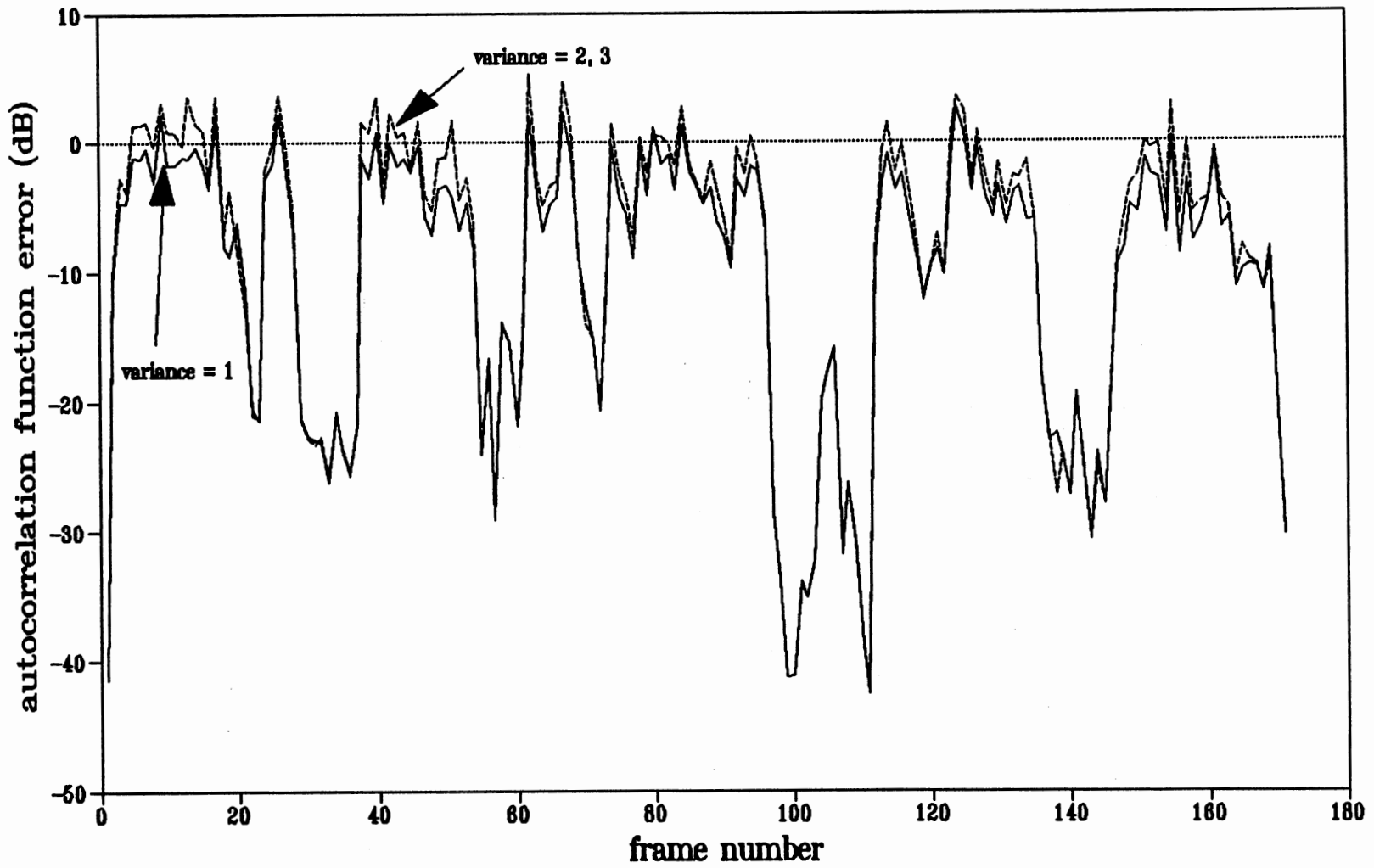


Figure 5.23 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different Noise Variances.

filter order is varied between 2, 5, and 10. As in previous experiment, the first 7680 samples of the outputs of the optimal IS filter of order 2 and 10 are depicted in Figure 5.24 and 5.25, respectively. Compare Figure 5.24 and 5.25 with the original speech (Figure 5.14). There is no significant visual difference between the three. As a result, in Figure 5.26, we compare the autocorrelation function error as a function of frame number for all three FIR filter orders. We can see that, for any specific frame, as the filter order increases, the autocorrelation function error tends to decrease. This is because as the filter order increases, more autocorrelation function lags can be matched to the autocorrelation function of the original speech resulting in improvement of signal reconstruction.

Note that the performance of the optimal IS filter in experiment 2 is not as good as in experiment 1. The result is not surprising since the bandwidth of the speech signal used in experiment 2 is smaller (we remind that sentence 2 is pronounced by a female speaker). It is known that a signal whose bandwidth is smaller also has slower decaying rate of the autocorrelation function, i.e., the autocorrelation function dies out to zero at longer lags. Note that the optimal IS filter matches the autocorrelation function up to the filter order lags. Thus, for the same amount of autocorrelation function error, a narrowband signal requires larger filter order to match to autocorrelation function than the wideband signal. In other words, for the same filter order, operating the optimal IS filter in the narrowband signal tends to yield larger autocorrelation function error than in the wideband signal. However, based on these two experiment results, the optimal IS filter still outperforms the

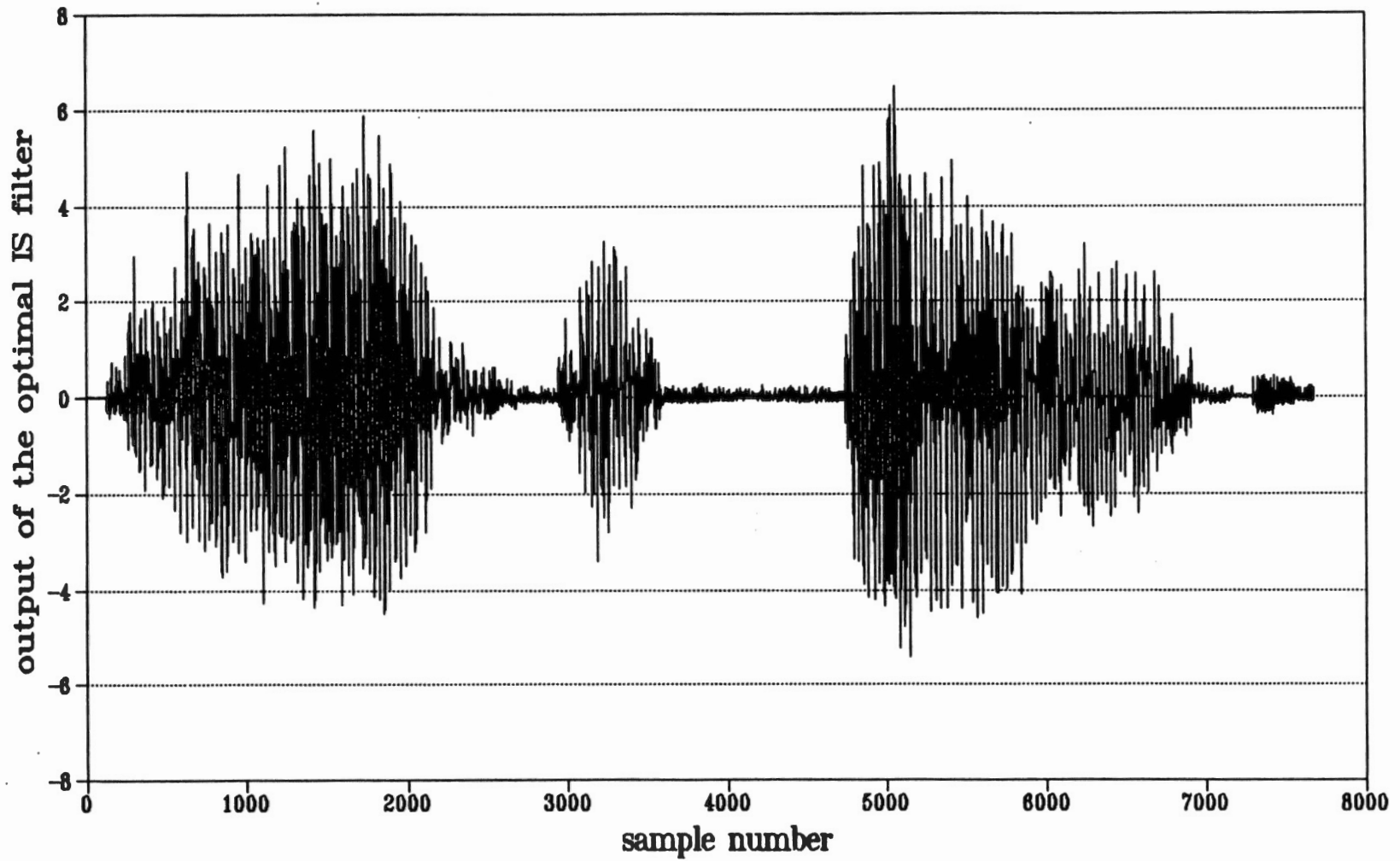


Figure 5.24 Plot of the First 7680 Samples of the Output of the Optimal IS Filter (Order = 2).

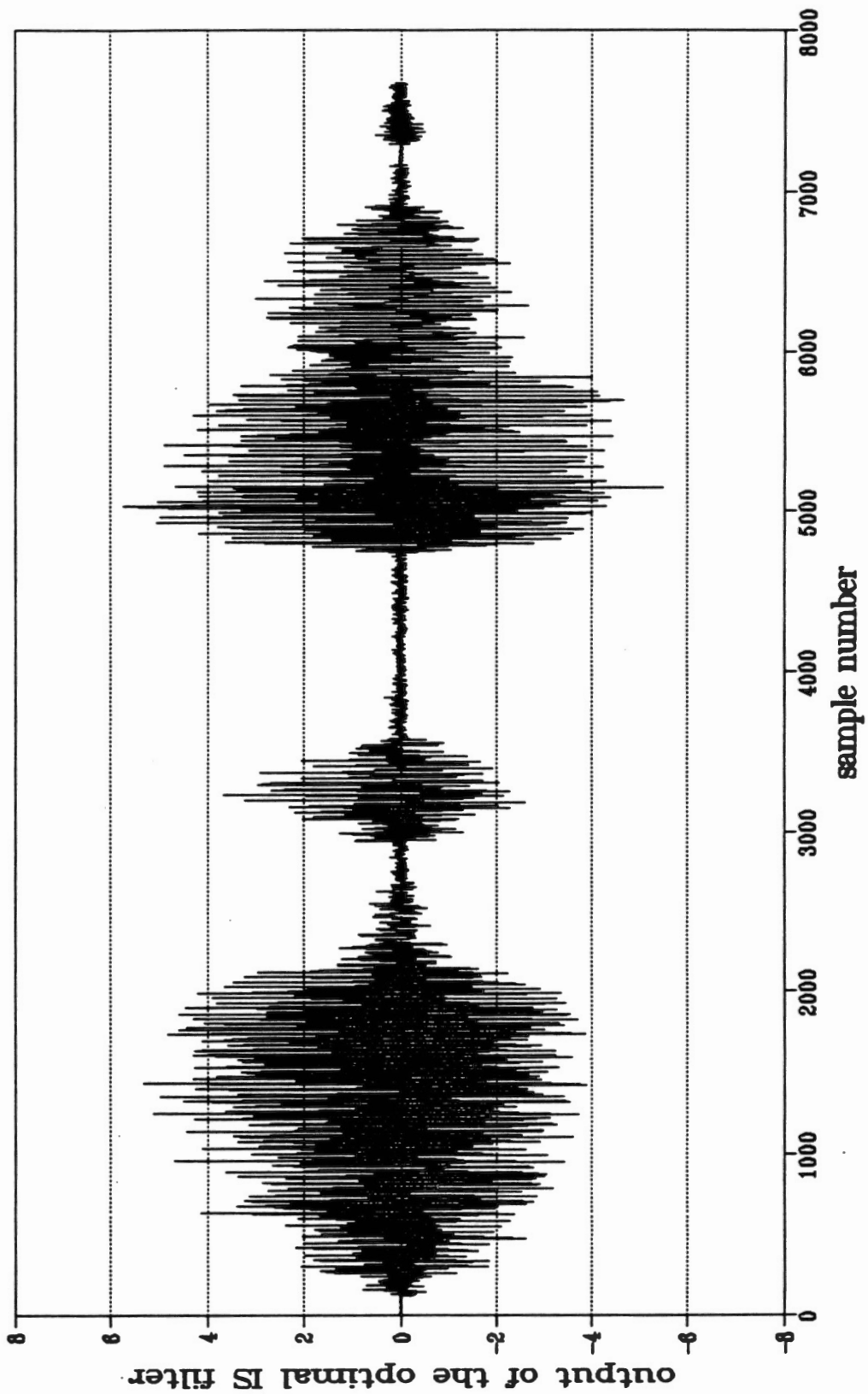


Figure 5.25 Plot of the First 7680 Samples of the Output of the Optimal IS Filter (Order = 10).

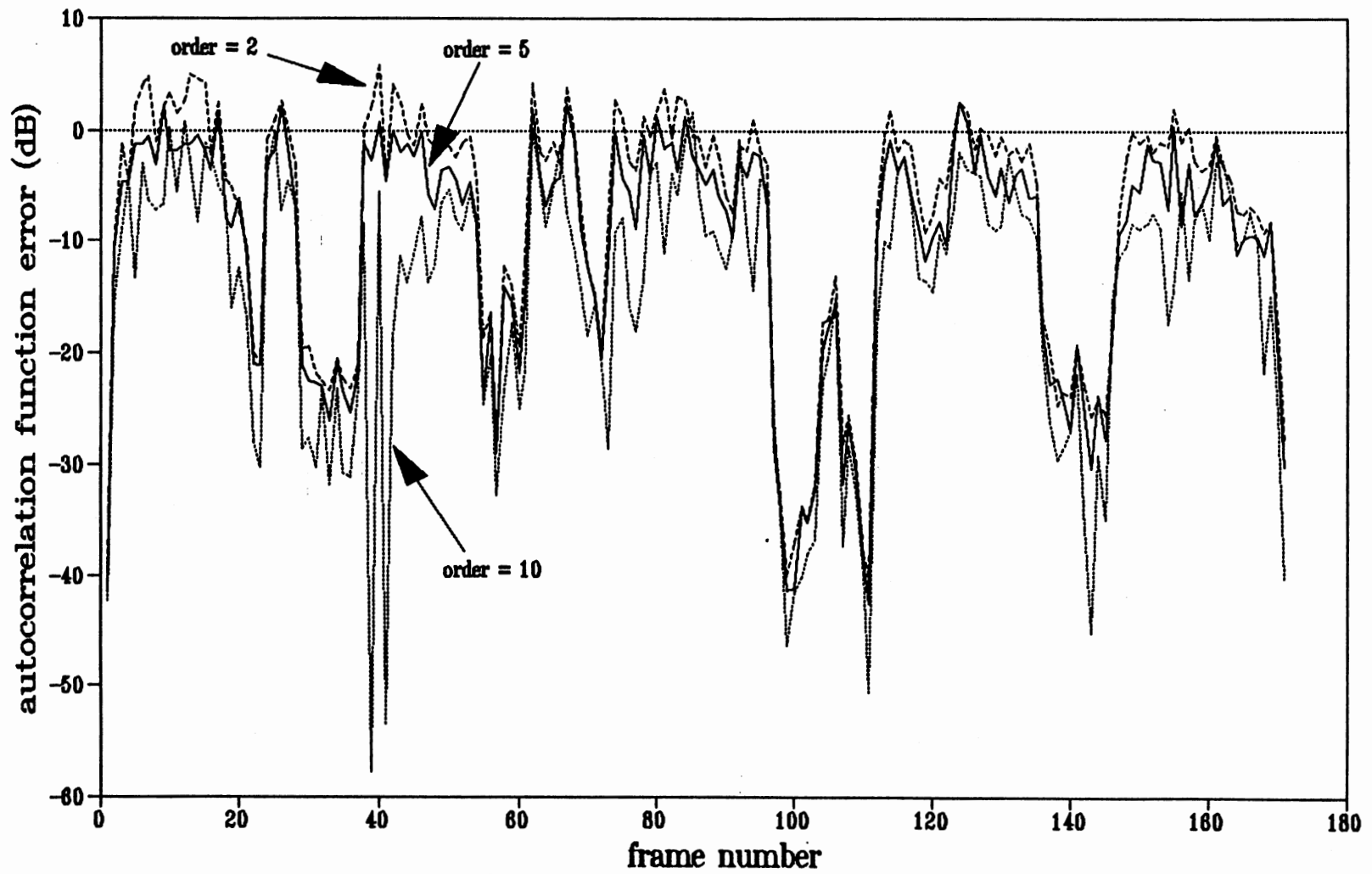


Figure 5.26 Comparison of the Autocorrelation Function Error of the Optimal IS Filter for Three Different Filter Order.

Wiener filter, in terms of matching the autocorrelation function, no matter whether the voice signal is narrowband or wideband.

Listening tests of both experiments were also performed. We have observed the following results:

1. The speech sound produced by the output of the optimal IS filter is louder than the speech sound produced by the output of the Wiener filter. This implies that the output of the optimal IS filter has higher output signal power than the output of the Wiener filter.

2. As the noise variance increases, the speech sound produced by the output of the Wiener filter tends to fade out. This is because as the noise variance increases, more signal energy is lost by the Wiener filter. For the optimal IS filter, the loudness level of speech output remains practically the same which makes it more perceptually understandable. However, as the noise variance increases, we have experienced more warbling sound in the output speech. This may be explained as follows. We have learned that the optimal IS filter preserves the signal energy by matching the autocorrelation function. Recall that the corrupting noise is assumed to be additive white Gaussian in nature. Theoretically, as the noise variance increases, only the autocorrelation function at the zeroth lag, $R_{yy}(0)$, of the received signal is bigger while the rest of the autocorrelation function remains the same. Thus, as the noise variance increases, the optimal IS filter has to put more effort to match the autocorrelation function at the zeroth lag with the expense of reduction in autocorrelation function matching in the other autocorrelation function lags. Note that the autocorrelation

function can also be defined as

$$R_{xx}(\tau) = E\{x(n)x(n+\tau)\} . \quad (5.4)$$

Thus, $R_{xx}(\tau)$ tells us the correlation between $x(n)$ and $x(n+\tau)$. In other words, for $\tau \neq 0$, $R_{xx}(\tau)$ contains the time localization information of the signal. Thus, the increasing of the noise variance causes more loss in signal time localization information, resulting in the warbling sound in the reproduced speech.

3. The warbling effect is more noticeable in experiment 2 than in experiment 1. This is reasonable since sentence two used in experiment 2 has narrower bandwidth which implies slower decaying rate in autocorrelation function. Thus, for the same performance, the experiment two requires larger filter order or larger input SNR than the experiment one.

4. Both Wiener filter and the optimal IS filter show perceptual improvement as the filter order increases. This is expected since larger order implies more autocorrelation function lags can be matched.

5.1.2 Jointly optimal pre- and post filter experiment

In this section, we also perform real speech simulation of jointly optimal pre- and post filter design case. The simulation will be the same as in Chapter 4, Figure 4.1. For simulation purposes, we restrict the FIR filter order of both pre- and post-optimal filter to be 5. Two experiments were performed. In the first experiment, sentence one is corrupted by the input noise and the channel noise as described in Chapter 4. Both corrupting noise sequences are assumed to be additive white Gaussian, with known

equal variances, 1. The first 7680 samples of the output speech of the jointly optimal system is depicted in Figure 5.27. Since the sum of the variance of the input noise and variance of channel noise is equal to 2, this simulation result can be compared to the single optimal IS filter case where the corrupting noise variance is equal to 2 whose output speech is depicted in Figure 5.3. Comparing Figure 5.27 and 5.3, there is no significant visual difference in terms of signal resemblance to the original speech. As a result, in Figure 5.28, we compare the autocorrelation function error of the jointly optimal system and that of the single optimal IS post-filter as a function of frame number. We can see that, for any specific frame, the autocorrelation function error of the jointly optimal system is always less than the autocorrelation function error of the single optimal IS post-filter. This implies that by using the jointly optimal system, the autocorrelation function matching is improved; thus, the output speech should more closely resemble the original speech.

In the second experiment, sentence two is also corrupted by both input and channel noise sequences which are assumed to be additive white Gaussian distributed of equal variances, 1. The simulation result is then compared with the single optimal IS filter case where the corrupting noise variance is equal to 2 (Note that, in the case, the output speech is depicted in Figure 5.16). The first 7680 samples of the output speech of the jointly optimal system is shown in Figure 5.29. Compare Figure 5.29 with 5.16, again there is no significant visual difference in terms of signal resemblance to the original speech. Thus, in Figure 5.30, we plot the autocorrelation function error versus the frame number for both jointly optimal case and single optimal IS post-filter

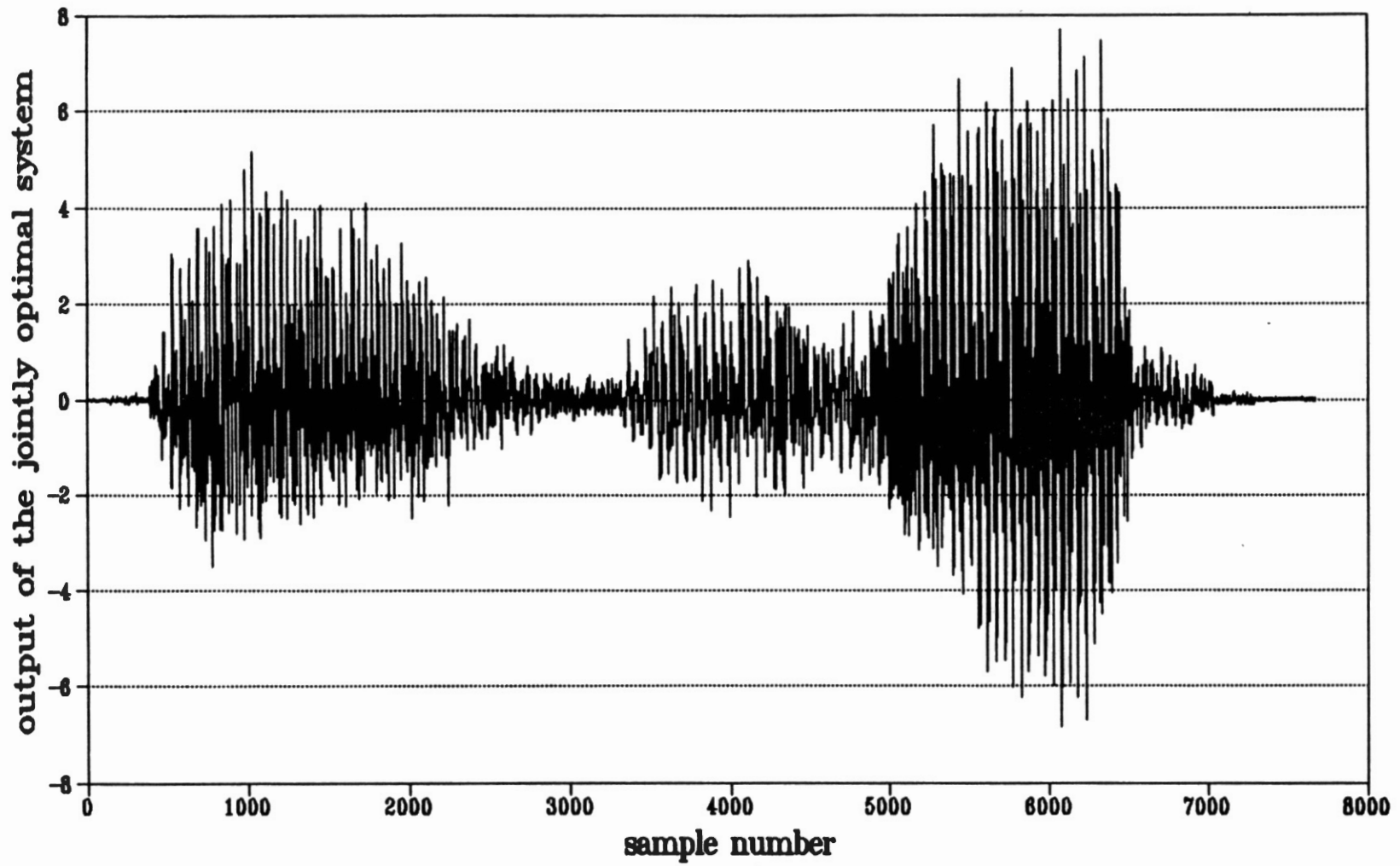


Figure 5.27 Plot of the First 7680 Samples of the Output of the Jointly Optimal System (Sentence One).

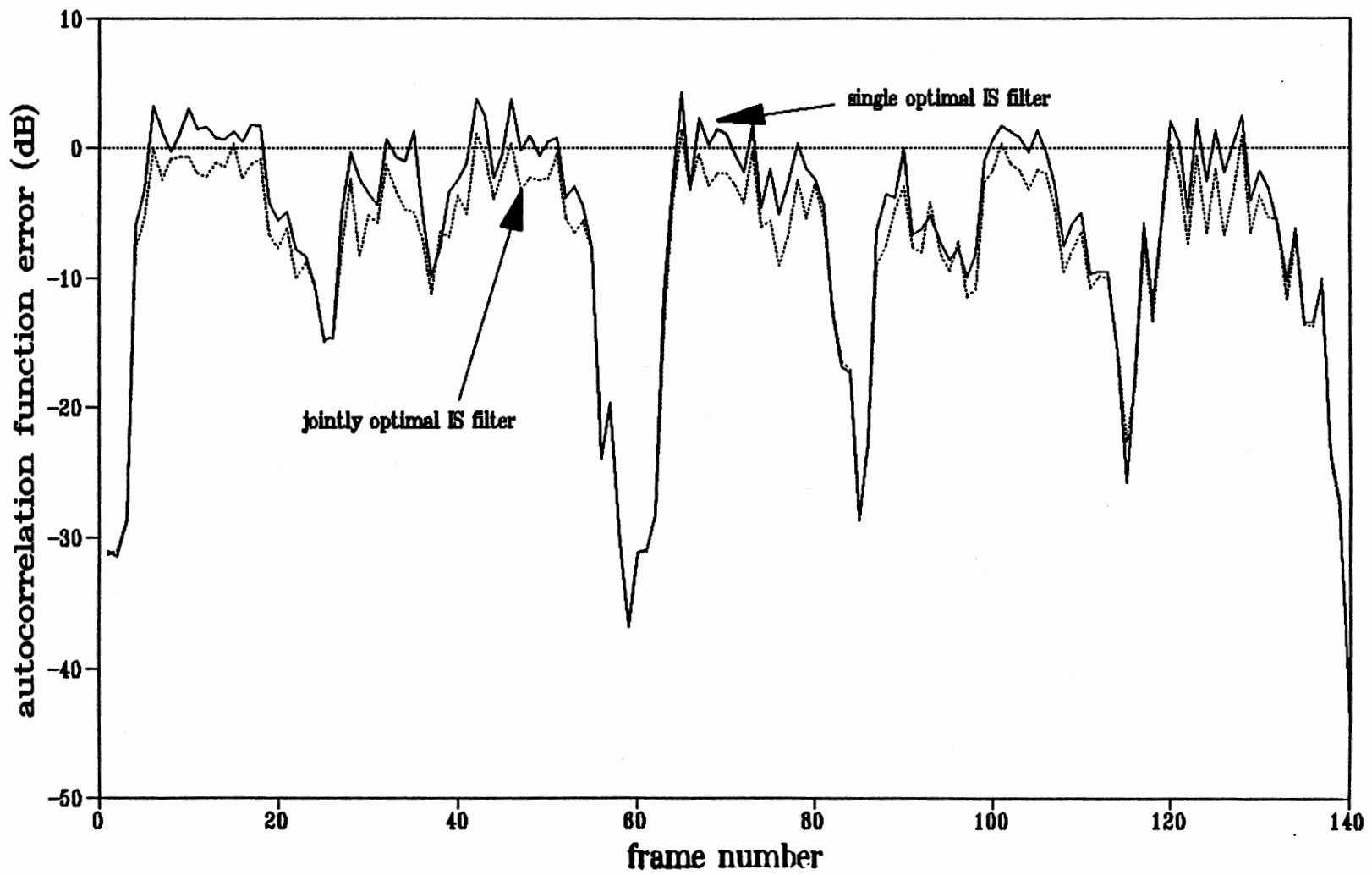


Figure 5.28 Comparison of the Jointly Optimal System and the Single Optimal System in Terms of the Autocorrelation Function Error (Sentence One).

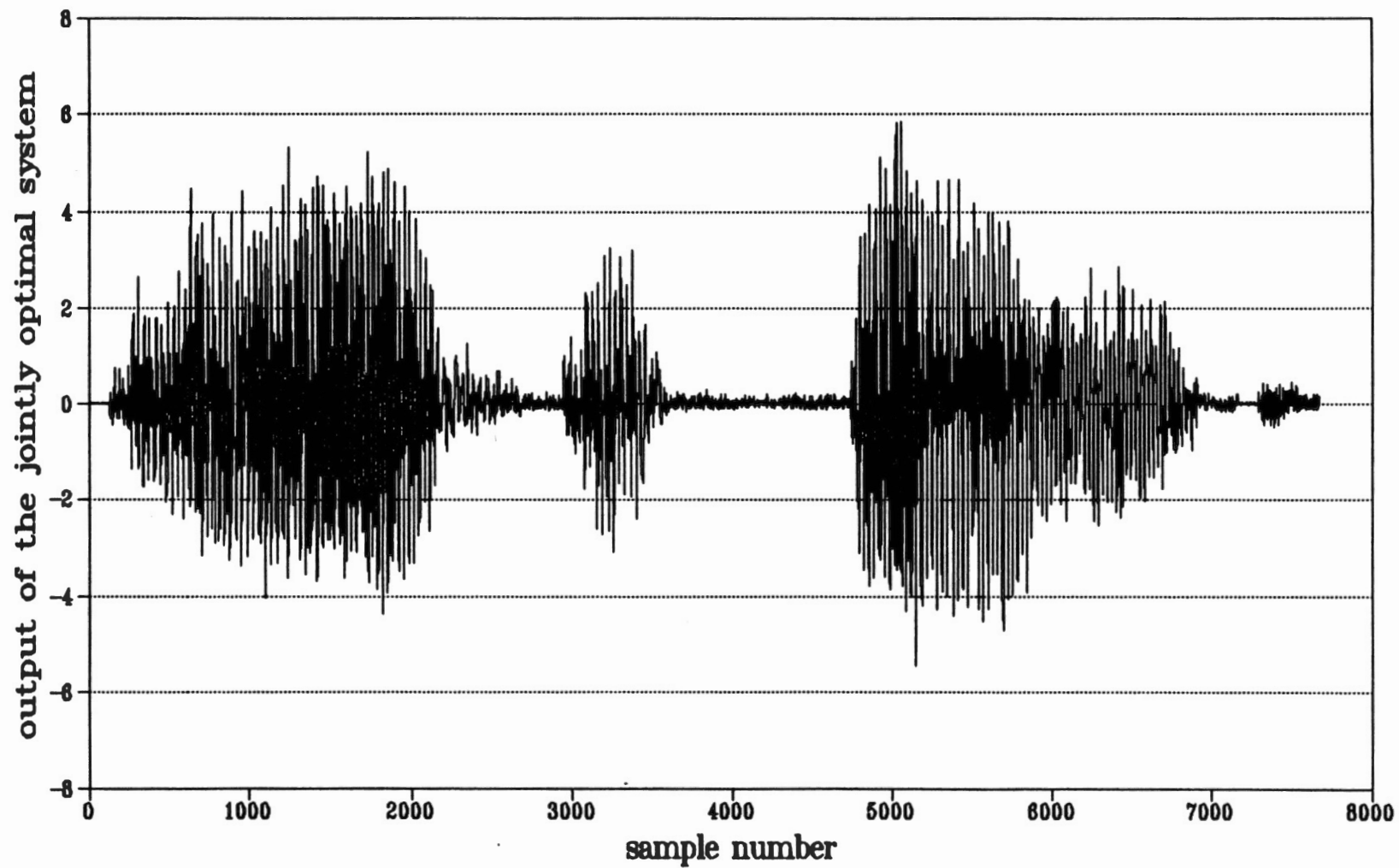


Figure 5.29 Plot of the First 7680 Samples of the Output of the Jointly Optimal System (Sentence Two).

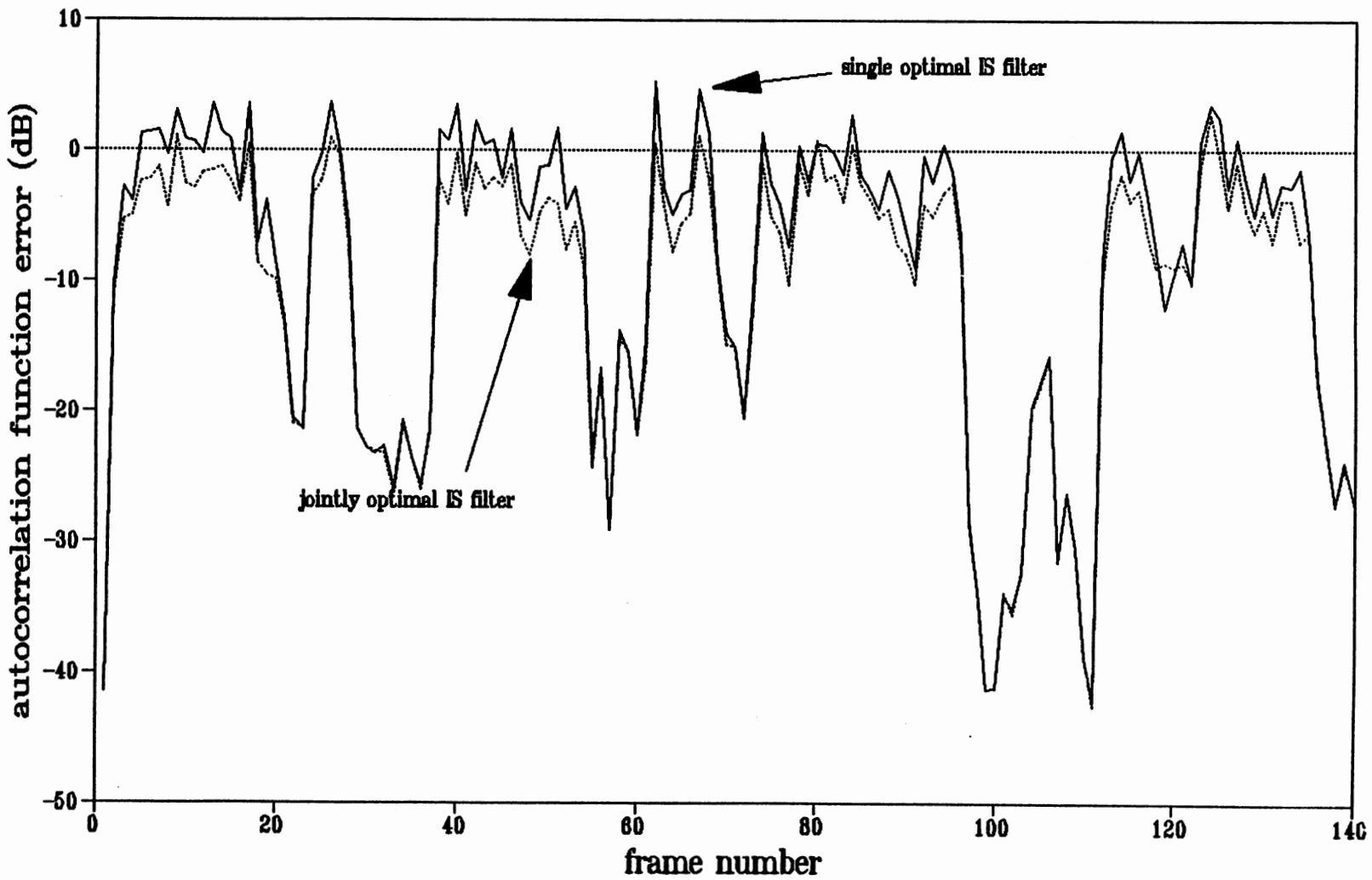


Figure 5.30 Comparison of the Jointly Optimal System and the Single Optimal System in Terms of the Autocorrelation Function Error (Sentence Two).

case. Again, for a specific frame, the autocorrelation function error of the jointly optimal case is usually less than the single optimal IS filter case which implies improvement in terms of autocorrelation function matching.

5.2 Summary

In this Chapter, we discussed simulation results that compared the performance of the Wiener filter and the optimal IS filter for real speech signals. The simulation results reveal that if the speech signal is corrupted by white Gaussian noise, the optimal IS filter outperforms the Wiener filter in terms of both minimizing IS distortion measure and the autocorrelation function matching, which agrees with the assumptions made in the previous Chapters.

CHAPTER VI

APPLICATION OF THE OPTIMAL IS FILTER IN DISCRETE COSINE TRANSFORM CODING

6.1 Introduction

All of the work presented to this point is based on the idea of trying to preserve signal portions which contain high energy levels. This is equivalent to preserving the mean and autocorrelation function matching property of the processed signal. However, the autocorrelation function does not convey the phase information. In other words, it does not provide the time localization contained in the phase information of the spectral components [Hla92]. As discussed in Chapter 3, even though the optimal covariance matrix is unique, the solution of the normal equations is not, which is due to the lack of phase information in the autocorrelation function. Note that for the high SNR environment (toll quality), the human auditory system is fairly insensitive to phase distortion. However, as the corrupting noise variance increases, more phase information is destroyed, causing this impairment to be more perceptually noticeable. As a result, the designing of a speech communication system under low SNR must take into account phase distortion. For instance, in the optimal IS filter case discussed in the Chapter 5, as the noise variance increases (the input

SNR goes down), the optimal IS filter tends to overfit the autocorrelation function at lower lags at the expense of underfitting the autocorrelation function at higher lags. Note that this impairment makes the processed sample more uncorrelated, causing a warbling effect noted in the listening tests.

One possible way to alleviate these problems is to code the signal into another orthogonal transform domain. Many orthogonal transforms have been successfully used in data compression applications. One of the most widely used orthogonal transforms is the discrete cosine transform (DCT) due to its near optimum performance with respect to variance distribution and its property of reducing block edge effects in image compression [Cla81]. In other words, the DCT has excellent energy compaction compared to the Fourier transform [Jai89]. As a result, within an allowable tolerance, a time domain sequence of length N can be represented by a DCT sequence of length much smaller than N , greatly reducing the data rate. Furthermore, the trend of using signal processing in the DCT domain is very promising due to the development of DCT integrated circuits (IC) by companies such as LSI Logic Corp. With this IC and additional coding algorithms, video data at a 100:1 compression ratio with close to analog videotape quality has been reported [Ang91].

The simplest schematic of DCT coding system is depicted in Figure 6.1. A frame of input signal, $x(n)$, of length N is first transformed into the DCT domain to obtain a sequence $v(n)$. Then, the last K samples of the sequence $v(n)$ are discarded. The transmitter transmits the rest of $N-K$ samples of $v(n)$ through the communication channel. At the receiver, the received sequence, $r(n)$, is the sum of the transmitted

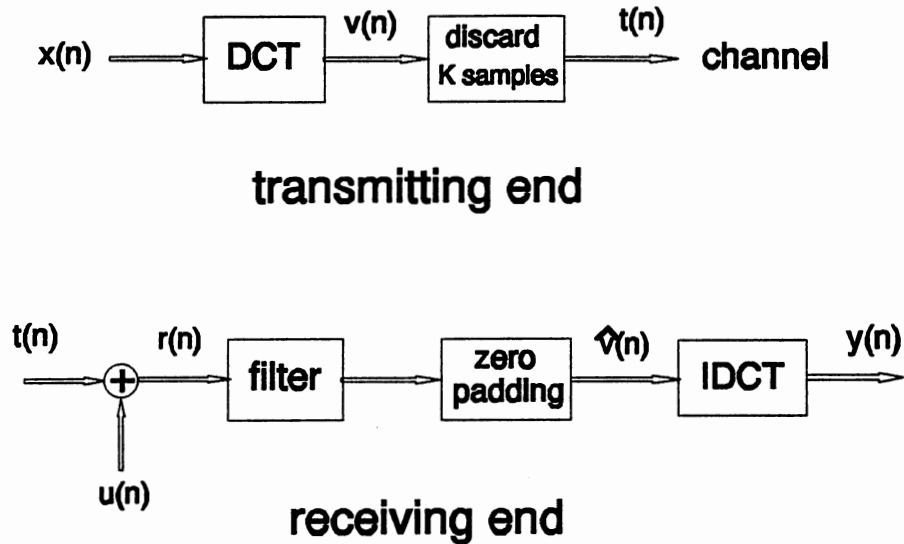


Figure 6.1 DCT Communication System.

sequence, $v(n)$, and the corrupting noise sequence, $u(n)$. The received signal, $r(n)$, is then filtered, padded with K samples of zeros, and transformed back to time domain to obtain the estimated signal, $y(n)$.

The communication system depicted in Figure 6.1 is also known as a DCT coder. Currently, all of transform coders are designed under an assumption of lossless criteria (noise free environment). The goal is to find a way to perfectly reconstruct the signal after the transform coefficients are critically decimated, i.e., the decimation rate is equal to the inverse of the number of FIR filter in a filter bank [Vai87, Vai89, Vai90]. Unfortunately, this noise free environment does not exist in practice.

One may argue that we may use any existing filter such as the Wiener filter to reduce the corrupting noise in the communication system. The first purpose of this

Chapter is to show that under the DCT domain, the optimal IS filter outperforms the Wiener filter in terms of preserving signal energy and minimizing IS distortion measure, making it more recognizable. The second purpose of this Chapter is to compare the performance of the optimal IS filter operating in the DCT domain with the performance of the optimal IS filter operating in time domain. We will show that the optimal IS filter performs better in the DCT domain than in the time domain in terms of minimizing the IS distortion measure. Furthermore, by operating in the DCT environment, the warbling effect caused by phase distortion can be reduced, which makes the processed signal more perceptually attractive.

6.2 Optimal IS filter in the DCT Domain

The DCT of a time domain sequence $x(n)$ of length N , $v(k)$, can be expressed as [Ahm74]

$$v_x(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos \left[\frac{\pi(2n+2)k}{2N} \right], \quad 0 \leq k \leq N-1 \quad (6.1)$$

where

$$\alpha(0) = \sqrt{\frac{1}{N}} \quad (6.2)$$

and

$$\alpha(k) = \sqrt{\frac{2}{N}}, \quad 1 \leq k \leq N-1. \quad (6.3)$$

The inverse discrete cosine transform (IDCT) is defined as

$$u(n) = \sum_{k=0}^{N-1} \alpha(k)v(k)\cos\left[\frac{\pi(2n+1)k}{2N}\right], \quad 0 \leq n \leq N-1. \quad (6.4)$$

We note that equation (6.1) can also be written in matrix form as

$$\mathbf{V} = \mathbf{A}\mathbf{X}, \quad (6.5)$$

where

$$\mathbf{V} = [v(0) \ v(1) \ \dots \ v(N-1)]^T, \quad (6.6)$$

$$\mathbf{X} = [x(0) \ x(1) \ \dots \ x(N-1)]^T, \quad (6.7)$$

and

$$\mathbf{A} = \begin{bmatrix} \alpha(0) & \dots & \alpha(N-1) \\ \alpha(0)\cos\left[\frac{\pi}{2N}\right] & \dots & \alpha(N-1)\cos\left[\frac{\pi(2N-1)}{2N}\right] \\ \vdots & \ddots & \vdots \\ \alpha(0)\cos\left[\frac{\pi(N-1)}{2N}\right] & \dots & \alpha(N-1)\cos\left[\frac{\pi(2N-1)(N-1)}{2N}\right] \end{bmatrix}. \quad (6.8)$$

Note that the DCT is real which makes it more attractive in terms of computation load. Furthermore, the DCT is orthogonal and unitary. Thus, there is no phase information loss during the transformation process.

We now define μ_x and R_x as the mean vector and the covariance matrix of the

$x(i)$ s. Let μ_v and R_v be the mean vector and covariance matrix of the $v(i)$ s. Since the DCT is unitary, we can write [Jai89]

$$\mu_x = A\mu_v, \quad (6.9)$$

and

$$R_v = AR_xA^T. \quad (6.10)$$

It is well known that the DCT has the property of packing a large fraction of the average energy of the input signal, $x(n)$, into a relatively few components of the transform coefficients, $v(n)$. This means that compared with R_x , the off-diagonal terms of R_v tend to become small compared to the diagonal elements, resulting in improvement of energy compaction.

However, the reduction of the bit rate comes with the price of larger bandwidth and loss of signal energy. Note that in the conventional communication system, the time domain signal, $x(n)$, is transmitted through the communication channel while in the DCT coding system, the DCT coefficients $v(n)$ are transmitted through the communication channel instead. Compared with $x(n)$, $v(n)$ is more impulse like (whiter). Thus, transmitting $v(n)$ through a communication channel may require larger bandwidth than transmitting $x(n)$ through the communication channel. Furthermore, note that even though the value of the last K samples in the DCT coefficients is relatively small, there is still a small amount of signal energy associated with them. By throwing away the last K samples of the transformed coefficients, a fraction of the signal energy will be lost. Thus, the larger the number of DCT coefficients discarded,

the more robust the communication system is required to be regarding distortion.

It is known that the Wiener filter will perform well for a narrowband signal whereas its efficiency degrades considerably as the signal bandwidth increases. This is because the Wiener filter is derived based on minimizing the MSE. As mentioned in Chapter 2, computing the MSE is based on summing the square of the difference between the original signal and the estimated signal. As a result, every signal sample, regardless of how much energy it possesses, has equal contribution to compute the MSE. However, in transform coding, the accuracy the first few samples of the transform coefficients is more critical since they contain the majority of the signal energy. As a result, filters used in a transform coding system should weigh the received samples according to their energy levels instead of the MSE. Thus, use of the Wiener filter under the DCT domain environment is not a very attractive solution.

It was shown in the previous Chapters that the optimal IS filter preserves the autocorrelation function matching property. This is equivalent to preserving the high energy portion of the input signal which is suitable to the DCT domain application. Thus, one would expect that compared to the Wiener filter, the optimal IS filter will perform better under the DCT environment.

Note that $v(n)$ has better energy compaction than $x(n)$; as a result, $V(f)$, the Fourier transform of $v(n)$, has larger bandwidth than $X(f)$, the Fourier transform of $x(n)$. Thus, more signal energy is compacted into fewer autocorrelation functions of $v(n)$ than the autocorrelation functions of $x(n)$. Since the optimal IS filter is simply matching the autocorrelation function, for the same filter order, operating the optimal

IS filter in the DCT domain should perform better than operating the optimal IS filter in the time domain in terms of preserving signal energy and minimizing the IS distortion measure.

6.3 Simulation Results and Discussions

In this section, we perform computer simulations to show how well the optimal IS filter operates under the DCT environment. The simulation scheme was the same as in Figure 6.1. Sentence one used in Chapter 5 was selected for experiments. The speech sentence is corrupted by an additive white Gaussian noise of known variance. The computer simulations were performed in two phases.

In the first phase, we compared the performance of the optimal IS filter with the Wiener filter in the DCT domain. Two corrupting noise variances are selected, 1 and 2. In this experiment, we assume that the time domain speech frame size N is equal to the DCT frame size $NDCT$ i.e., no DCT coefficients were discarded, $K = 0$. In addition, the frame size used in this experiment was 128, i.e., $N = NDCT = 128$, and the filter order was restricted to be 5. Figure 6.2 and 6.3 show the first 7680 samples of the IDCT of the output of the Wiener filter and the optimal IS filter for the case of corrupting noise variance equal to 1, respectively. Compare Figure 6.2 and 6.3, the output of the Wiener filter shows more loss of signal energy especially from sample 500 to sample 3000. In Figure 6.4 and 6.5, we compare the IS distortion measure as a function of the frame number of both the Wiener filter and the optimal IS filter in the case where the corrupting noise variance is varied from 1 and 2,

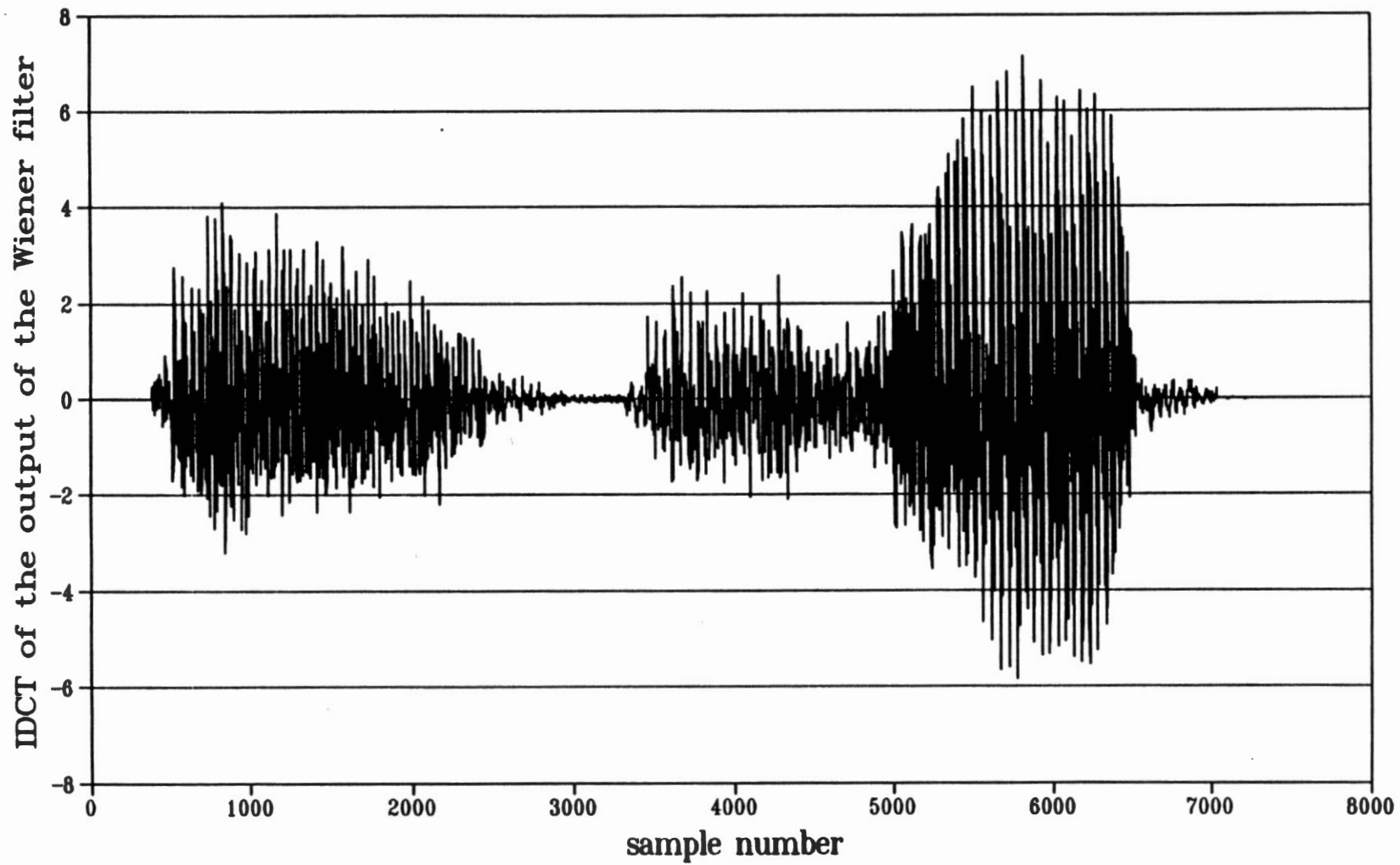


Figure 6.2 Plot of the First 7680 Samples of the IDCT of the Wiener Filter Output.

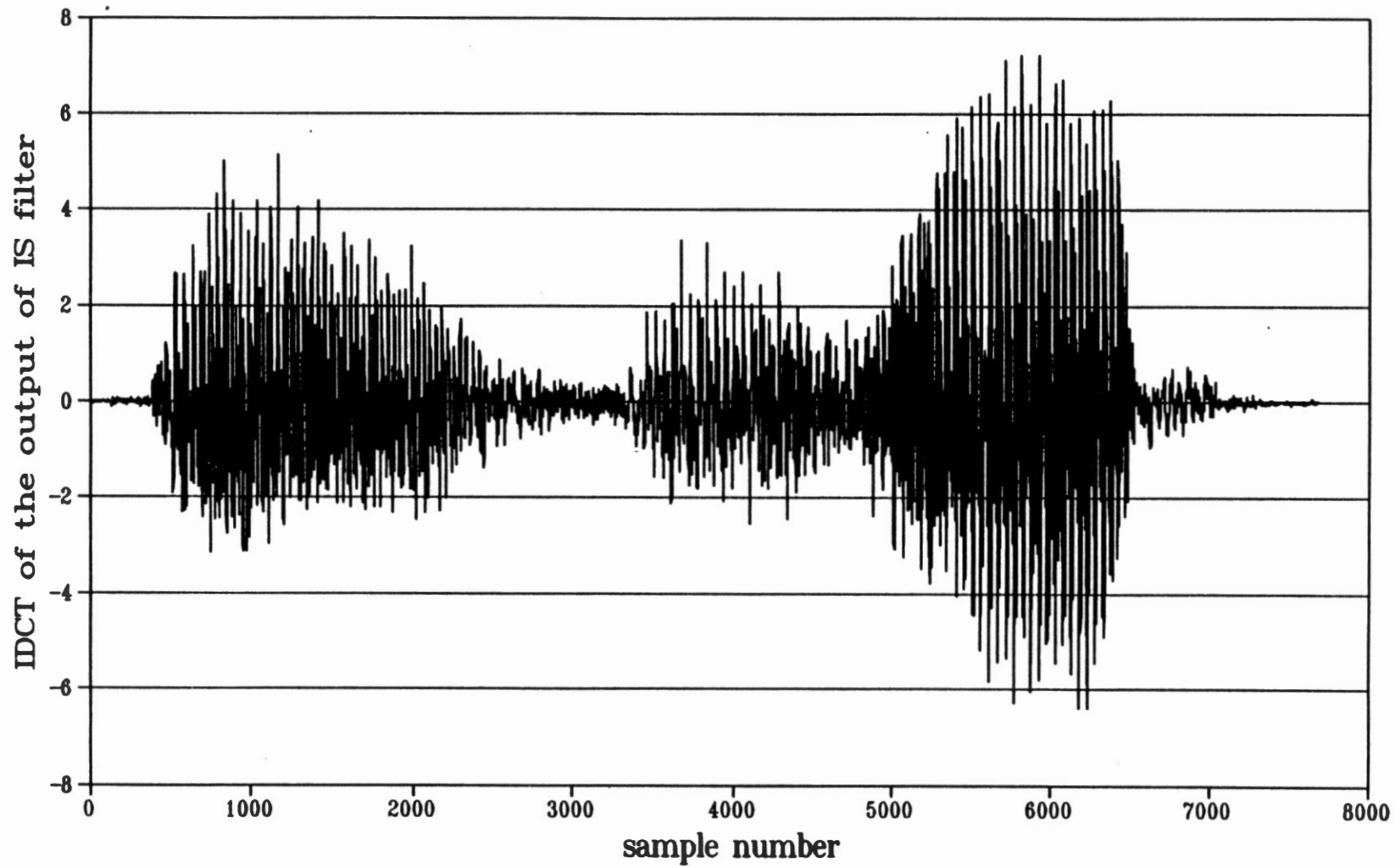


Figure 6.3 Plot of the First 7680 Samples of the IDCT of the Optimal IS Filter Output.

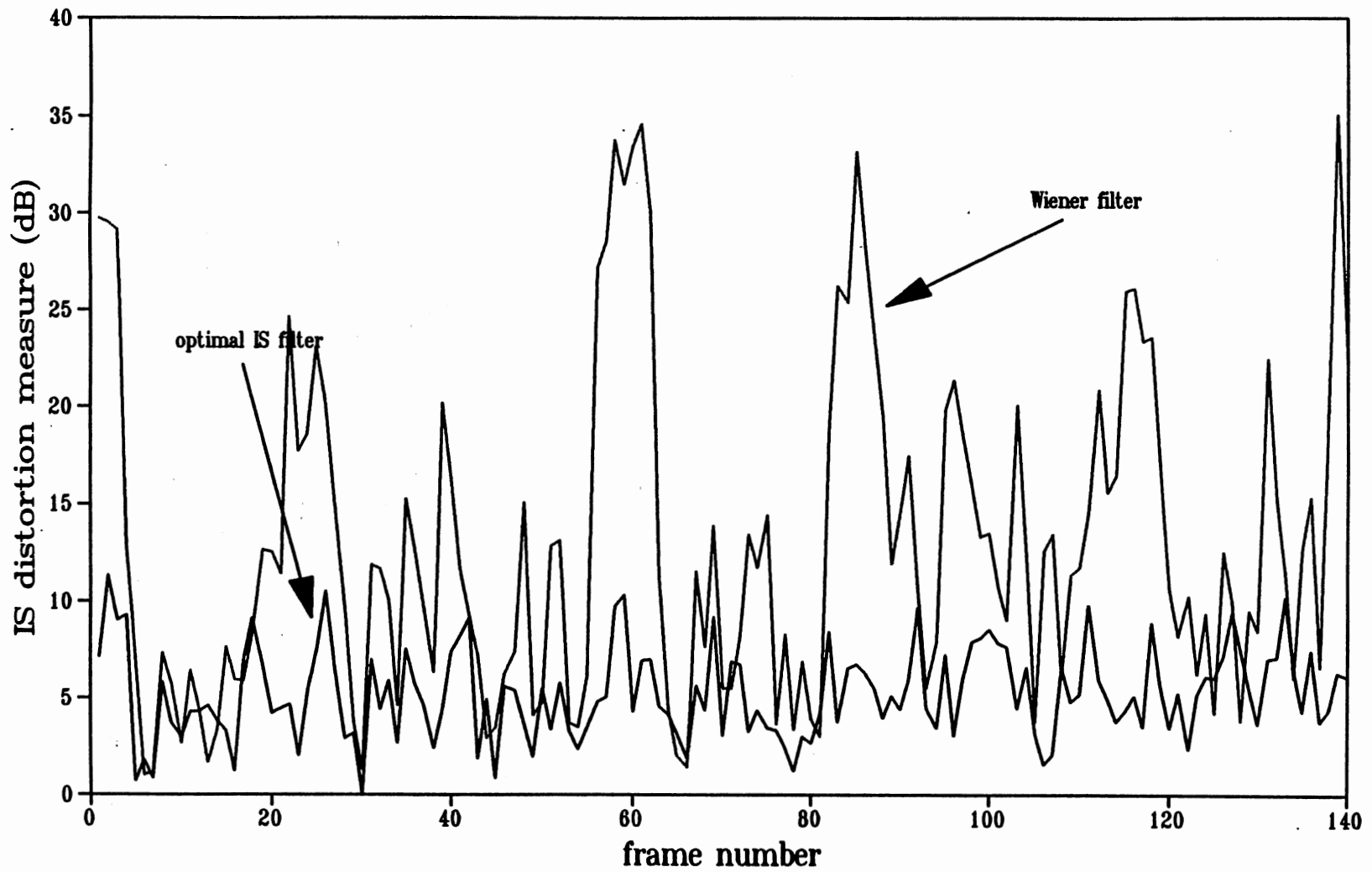


Figure 6.4 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 1$).

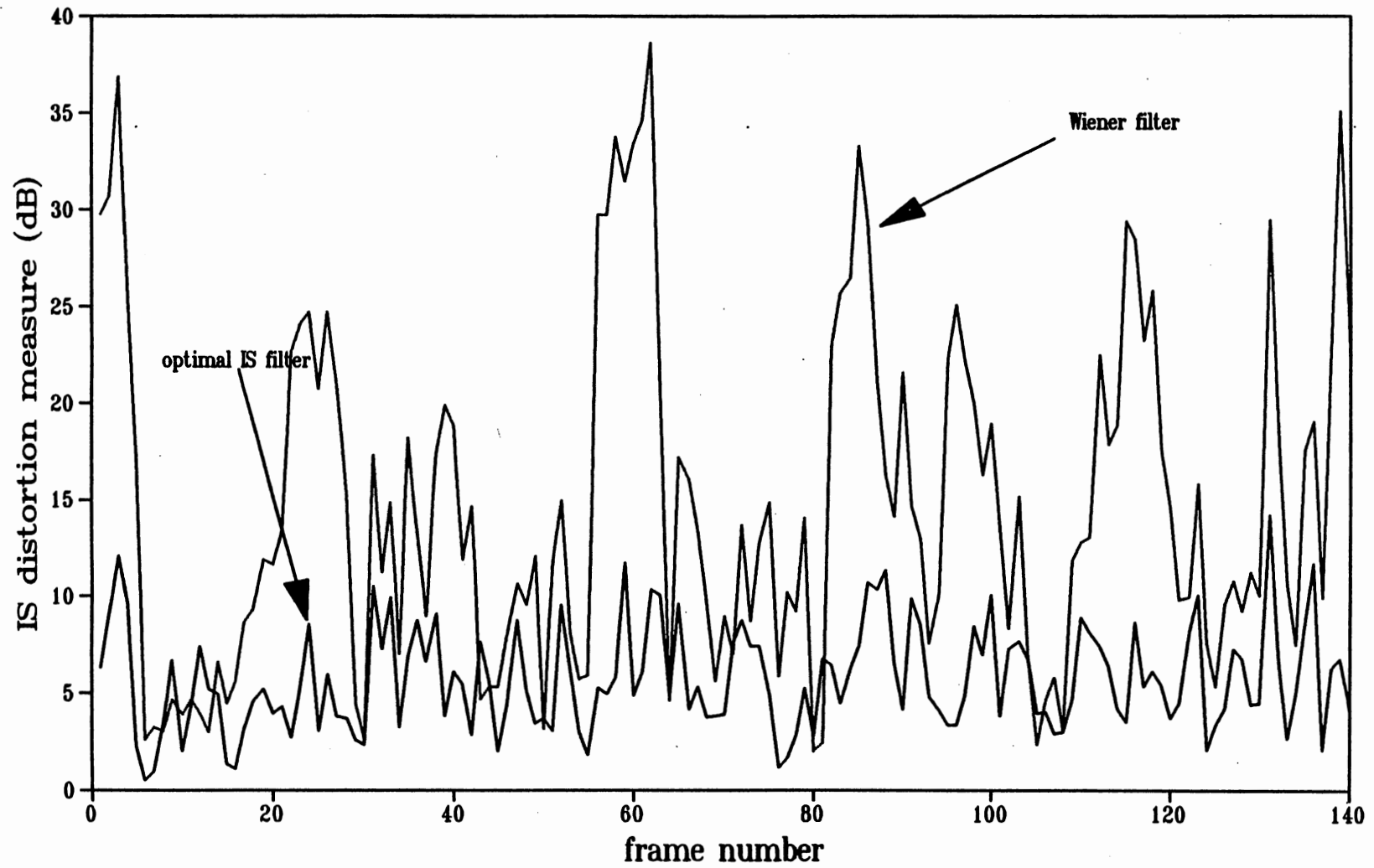


Figure 6.5 Comparison of the Wiener Filter and the Optimal IS Filter in Terms of the IS Distortion Measure ($\sigma^2 = 2$).

respectively. From Figure 6.4 and 6.5, we can see that the IS distortion measure of the optimal IS filter is usually well below the IS distortion measure of the Wiener filter. Notice that for some specific frames, e.g., frame 55 to frame 65, the Wiener filter yields very large IS distortion measure. Thus, we conclude that under the DCT environment, the optimal IS filter outperforms the Wiener filter in terms of minimizing the IS distortion measure.

We then investigate the performance of the optimal IS filter in the DCT coding system where the last 32 samples of each frame is discarded, i.e, $K = 32$. The corrupting noise variance is assumed to be 1. Thus, the data rate is reduced by 25%. Figure 6.6 shows the first 7680 samples of the output of the optimal IS filter for $K = 32$. Note that there is no distinctive different between Figure 6.6 and 6.3 in terms of signal resemblance to the original speech. However, in Figure 6.7, we compare the IS distortion measure of the optimal IS filter for $K = 32$ and $K = 0$ as a function of frame number. From Figure 6.7, the IS distortion measure for the case where $K = 32$ tends to be more than the IS distortion measure for the case where $K = 0$. However, compared to the Wiener filter (Figure 6.4 and 6.5), the IS distortion measure of the case where $K = 32$ is still smaller despite the fact that the data rate is reduced by 25 %. Thus, compared to a conventional speech communication system, where the time domain signal is transmitted through the communication system cooperating with the Wiener filter, use of the optimal IS filter in the DCT domain yields several advantages.

1. The IDCT of the output of the optimal IS filter in the DCT domain yields

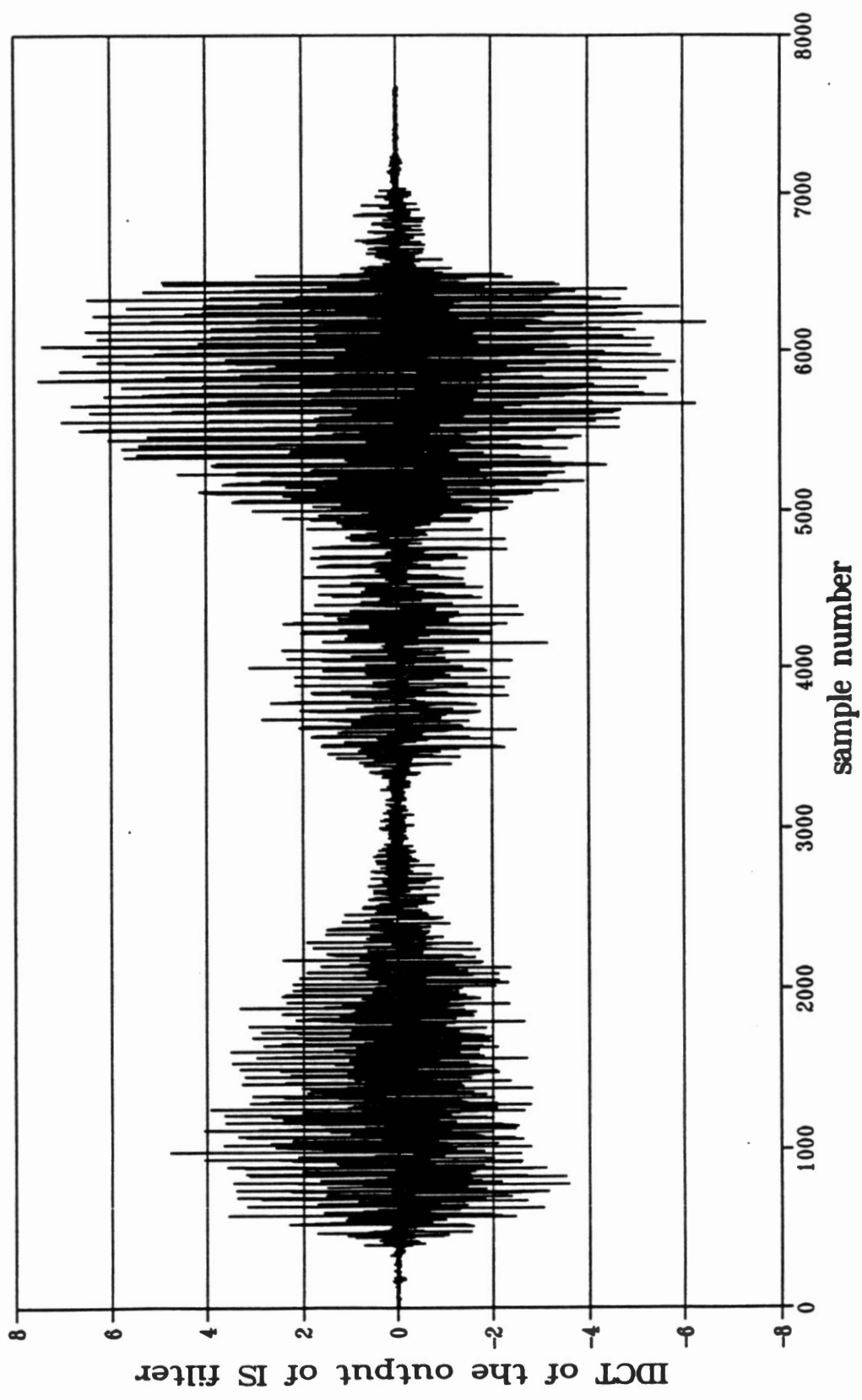


Figure 6.6 Plot of the First 7680 Samples of the Optimal IS Filter ($K = 32$).

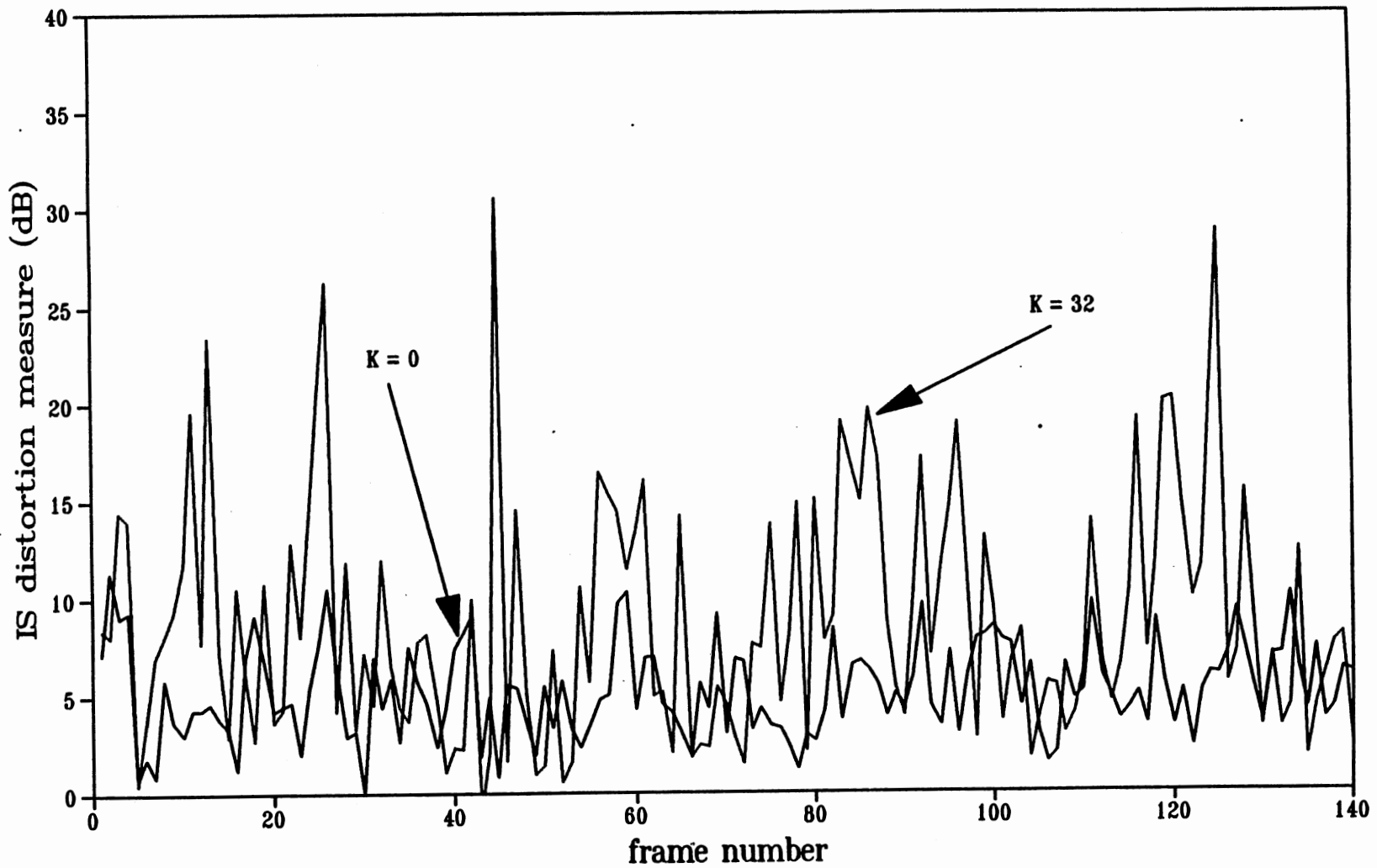


Figure 6.7 Comparison of the IS Distortion Measure of the Optimal IS Filter for $K = 0$ and $K = 32$.

lower IS distortion measure which makes it more perceptually desirable.

2. The autocorrelation function of the IDCT of the output of the optimal IS filter in the DCT domain matches better to the autocorrelation function of the original speech. Thus, more signal energy is preserved.

3. Since the DCT has excellent energy compaction, the data rate can be reduced with a small amount of signal degradation.

The second phase of the computer simulation is to compare the performance of the optimal IS filter in the DCT domain with that of the optimal IS filter in the time domain. Figure 6.8 and 6.9 compare the IS distortion measure of both the optimal IS filter in the time domain and the optimal IS filter in time domain for the case where the noise variance is 1 and 2, respectively. Note that the filter order is restricted to be 5. From Figure 6.8 and 6.9, under the DCT domain, the IS distortion measure of the optimal IS filter is smoother and well below 16 dB. However, in the time domain, the IS distortion measure of the optimal IS filter is very large for some speech frames. Note that a speech frame of large IS distortion measure is much more perceptually noticeable than a speech frame of small IS distortion measure. Thus, the optimal IS filter performs better in the DCT domain than in the time domain in terms of minimizing the IS distortion measure.

Listening tests were also performed. We have observed some results as follows;

1. The reproduced speech of the IDCT of the output of the optimal IS filter is louder than the reproduced speech of the IDCT of the output of the Wiener filter. As

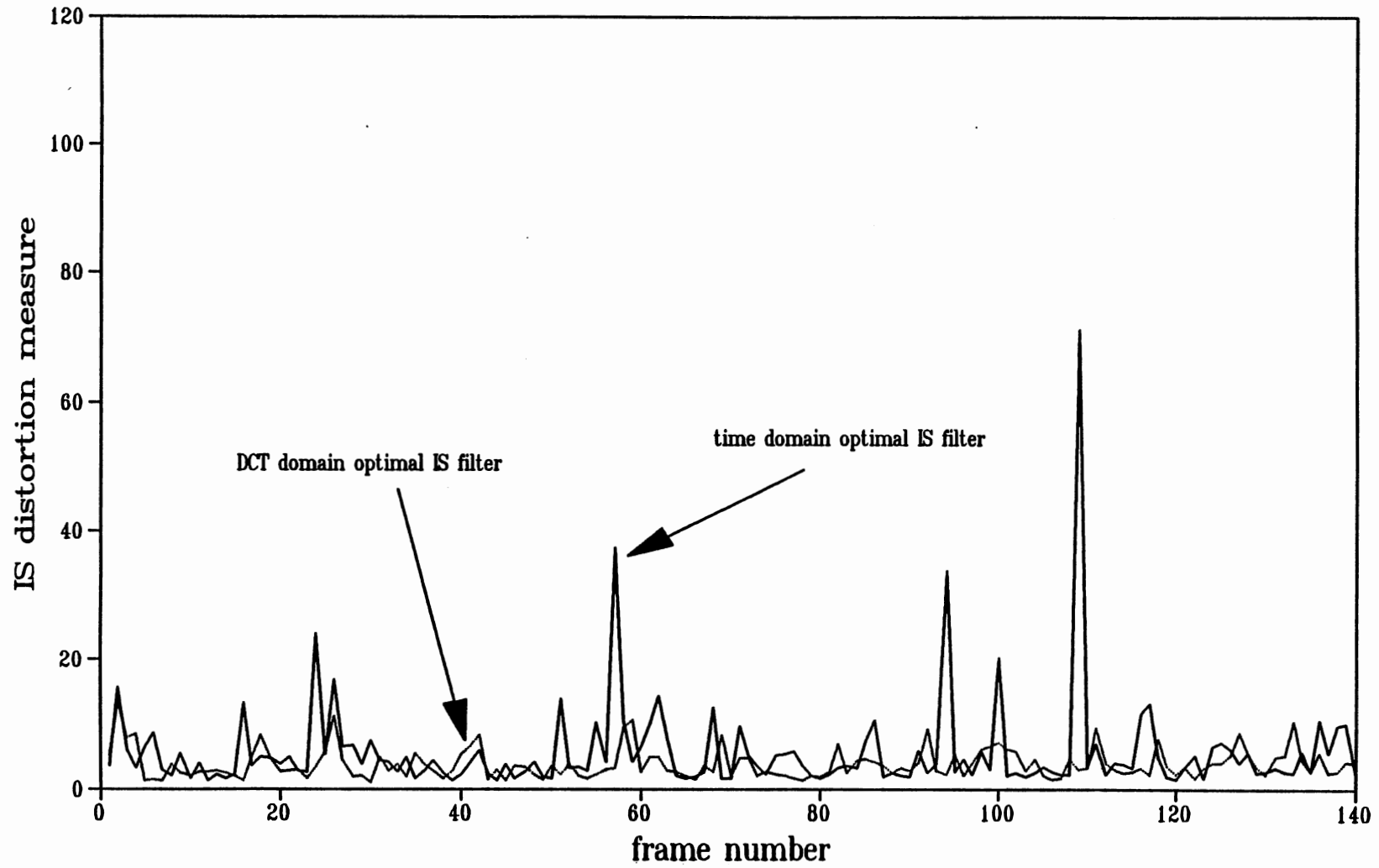


Figure 6.8 Comparison of the IS Distortion Measure of the Optimal IS Filter in the DCT Domain and Optimal IS Filter in the Time Domain ($\sigma^2 = 1$).

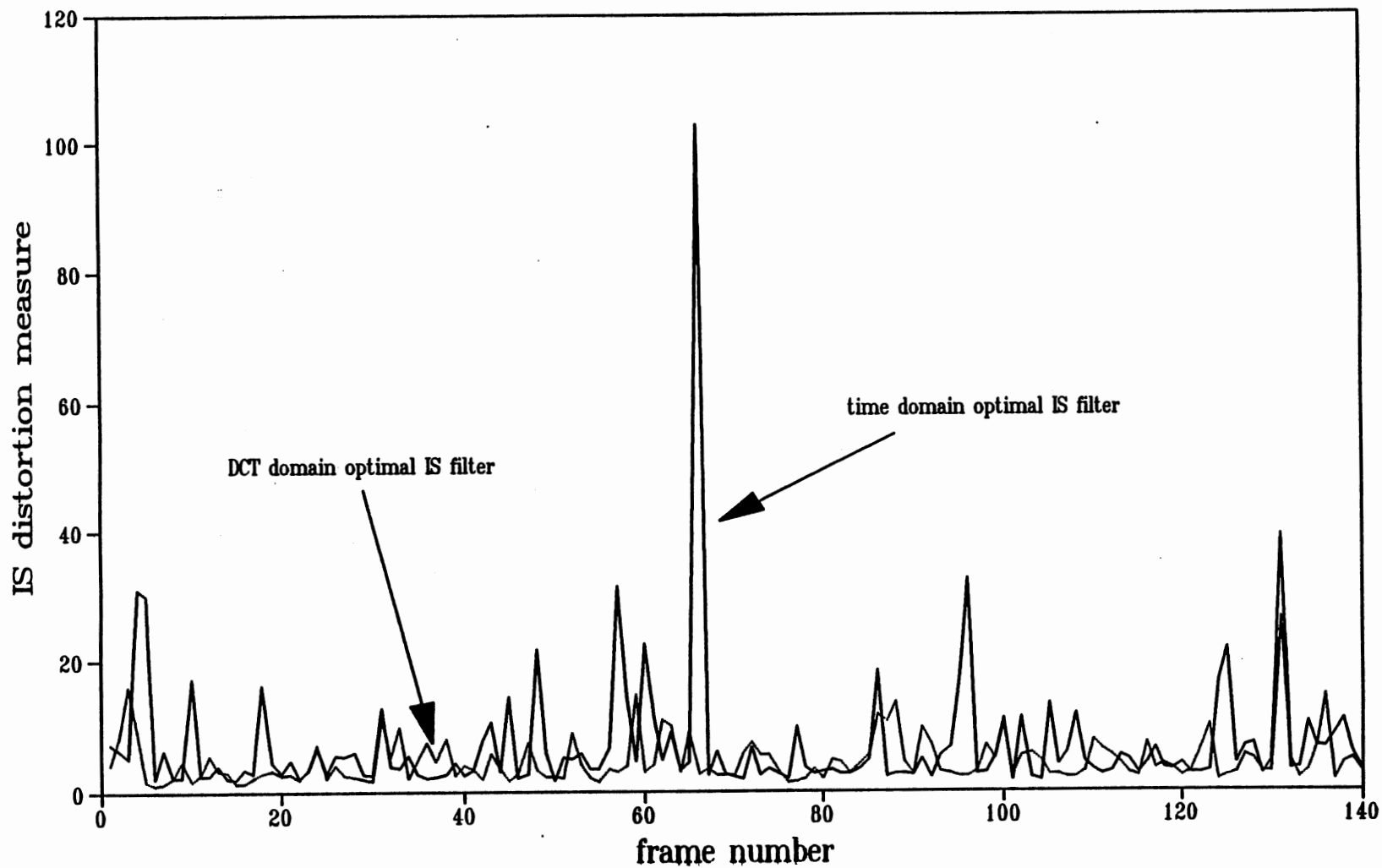


Figure 6.9 Comparison of the IS Distortion Measure of the Optimal IS Filter in the DCT Domain and Optimal IS Filter in the Time Domain ($\sigma^2 = 2$).

the noise variance increases, the reproduced speech of the IDCT of the output of the Wiener filter tends to fade out. The result is not surprising because as discussed before, more signal energy is lost by the Wiener filter as the noise variance increases. On the contrary, the loudness level of the IDCT of the output of the optimal IS filter still remains acceptable which makes it more easily recognizable. This is because the optimal IS filter preserves more signal energy than the Wiener filter.

2. We also compare the reproduced speech of the IDCT of the optimal IS filter in the DCT domain with the reproduced speech of the output of the optimal IS filter in the time domain. The warbling effect noted when operating the optimal IS filter in the DCT domain is much less than that observed when operating the optimal IS filter in the time domain. This is because the DCT coefficients are whiter than the original time domain signal, resulting in more compaction in autocorrelation function. Thus, perceptually, the optimal IS filter performs better in the DCT domain than in the time domain.

6.3 Summary

In this Chapter, we have shown that the optimal IS filter can be successfully used in an orthogonal transform coding system. The advantage of this type of coding system is the reduction of the data rate when the signal is transformed in such a way that the majority of the signal is packed in the first few transformed coefficients. The major drawback of this type of coding system is the increased bandwidth which makes it unsuitable to the Wiener filter application. Furthermore, the MSE depends on the

value of the difference between the original signal and the reconstructed signal regardless of how much energy it possesses. As a result, the more suitable filter in this application is the optimal IS filter which preserves the high energy level portion of the signal. Simulation results reveal that the optimal IS filter outperforms the Wiener filter in terms of minimizing the IS distortion measure which makes it more perceptually desirable. Furthermore, the optimal IS filter performs better in the DCT domain than in time domain since warbling sound is far less noticeable.

CHAPTER VII

CONCLUSIONS

7.1 Summary

In Chapter 1, we discussed the human speech production mechanism and the human auditory system. We noted that speech processing can be more easily performed in the frequency domain than in the time domain. We also noted that the first two formants played a major role in terms of speech recognition. As a result, we can categorize the speech sounds according to their phoneme group. Two speech sounds which belong to the same phoneme group will sound alike. Thus, for speech recognition purposes, a good receiver must be able to classify the received signal corrupted by any existing noise to its associated group.

The above technique is also known as the nearest neighborhood system in discriminant analysis. In discrimination analysis, the received speech signal will be assigned to the phoneme group which yields the smallest distortion measure. Several distortion measure can be used, for example, MSE, IS distortion measure, etc.. In Chapter 2, we discussed several types of distortion measures and briefly compared their advantages and disadvantages. We noted that MSE was not a good choice of a distortion measure for speech perception since a large MSE did not always imply poor

perceptual speech quality. We also concentrated on one special type of distortion measure called the IS distortion measure. We observed that the IS distortion measure was in fact twice the limit of the information discrimination function which was used as a tool to measure similarity between two Gaussian processes. Thus, for the nearest neighborhood system in speech perceptual point of view, the processed signal should yield the smallest IS distortion measure between the original signal and the processed signal.

As mentioned earlier, the first two formants play a dominant role in terms of recognizing which sound is produced. Thus, for perceptual purposes, the received signal should be processed in such a way that the first two formants are preserved. As a result, we proposed a strategy that a good perceptual speech signal processing algorithm should preserve the mean and autocorrelation function matching property, in turn, preserving the formant frequencies.

In optimal filtering, the filtered output is obtained by minimizing the distortion measure between the original signal, $x(n)$, and the estimated signal, $y(n)$. As a result, the optimal filtered output will yield the smallest distortion measure to the original signal from all possible output. Thus, it is possible to view optimal filtering technique as one of the discriminant analysis. Optimal filtering has played a major role in a communication systems for the past two decades. One of the most widely used distortion measures is the MSE between $x(n)$ and $y(n)$. The optimal filter which minimizes the MSE between $x(n)$ and $y(n)$ is called the Wiener filter. In Chapter 3, we showed that the Wiener filter did not preserve the autocorrelation function

matching property. Thus, from the speech perceptual viewpoint, the Wiener filter did not represent the nearest neighborhood system. We then proposed an alternate way to design a perceptually optimal FIR filter called the optimal IS filter. This optimal IS filter is obtained by minimizing the IS distortion measure between $x(n)$ and $y(n)$. We showed that this was equivalent to matching the frequency response of the Wiener filter with the magnitude square of the frequency response of the optimal IS filter. We then showed that the optimal IS filter did preserve the autocorrelation function matching property which made it more perceptually desirable. Some computer simulations were also performed to compare the performance of the Wiener filter with the optimal IS filter. Simulation results did agree with our theoretical derivation that the optimal IS filter outperformed the Wiener filter in terms of both spectral matching and the output SNR.

In Chapter 4, we improved the performance of the optimal IS filter using the jointly optimal pre- and post-filter design. With the use of a prefilter, the transmitting signal is changed into a form which is more robust to the existing noise in the communication system. The normal equations for the jointly optimal pre- and post-filter were derived. The suboptimal solution can be found via the use of Newton's algorithm. Computer simulation results showed that improvement could be made with the use of jointly optimal pre- and post-filter design in terms of both IS distortion measure and output SNR.

In Chapter 5, we performed computer simulations of the optimal IS filter on real speech signals. Two English sentences are selected for processing. Results

revealed that the optimal IS filter outperformed the Wiener filter in terms of minimizing the IS distortion measure and autocorrelation function matching. In the Wiener filter, as the input SNR decreases, more energy is lost by the Wiener filter causing the performance to degrade considerably. On the contrary, in the optimal IS filter, the loudness level of the optimal IS filter output remained acceptable which made it easier to be recognized. However, we experienced more warbling sound in the optimal IS filter output as the noise variance increased. The warbling was caused the overfitting the autocorrelation function in the lower lags at the expense of underfitting the autocorrelation function in the higher lags, causing more phase information to be loss.

In Chapter 6, we illustrated the application of the optimal IS filter in the DCT domain. It is known that the DCT increases the energy compaction of the time domain signal. In other words, the majority of the signal energy is packed in the first few transformed coefficients. Note that the DCT coefficients are whiter than the time domain signal. Thus, transmitting the DCT coefficients through a communication channel may require larger bandwidth than transmitting the time domain signal. In addition, it is known that the Wiener filter does not perform quite well in the wide band situation. This is because the MSE, which the Wiener filter minimizes, is computed based on the difference between the original signal and the estimated signal irregardless of how much energy the signal sample possesses. Recall that the basic goal of the optimal IS filter is simply matching the autocorrelation function. In addition, note that larger bandwidth implies that more energy is also packed in the first

few lags of the autocorrelation functions. Thus, the optimal IS filter performs much better in wideband environment such as the DCT domain.

Since the autocorrelation function of the DCT coefficients is more compact than that of the time domain signal, the optimal IS filter will perform better in the DCT domain than in the time domain. Simulation results also showed that the optimal IS filter in the DCT domain yields better performance than the optimal IS filter in the time domain in terms of minimizing the IS distortion measure. Furthermore, in listening tests the warbling sound of the output of the optimal IS filter in the DCT domain is far less noticeable than that of in the time domain.

7.2 Considerations for Future Research

There is still much research that can be done toward designing an optimal filter for speech signal processing. Up to the present, there is no optimal filter specifically designed for speech signal processing. This thesis serves as an introduction to the exploration of a this new area. However, speech quality involves subjective judgement, which no objective measurement can absolutely represent. Thus, the absolutely optimal speech signal processor still remains to be found.

We note that this thesis is limited to the FIR filter design. Further research can extend these results for designing an infinite impulse response (IIR) filter. We also assume that the signal is simply corrupted by white Gaussian noise of known variance. For the case of colored corrupting noise, some modifications will be needed in order to accurately estimate the autocorrelation of the speech signal.

During the listening test of the real speech, we experienced some warbling sound from the IS filter reconstructed speech. This problem arises from the fact that our algorithm is simply based on making use of speech redundancy and human hearing imperfections. Note that the IS distortion measure is basically computed from the power spectrum of the original speech and the processed speech where no phase information is taken into account. In other words, we are only matching the autocorrelation function and ignoring the phase information. For speech toll quality environment (high SNR), the human hearing system is fairly insensitive to phase variation. However, as the corrupting noise variance increases, the phase distortion becomes more noticeable and unbearable if the input SNR is low enough. Thus, to combat this type of distortion, the future speech distortion measure must be modified to include the phase information of the signal.

In Chapter 6, we explored a new area of digital filter design in the orthogonal transform domain. Future investigations in this area are worthwhile due to the rapid growth of VLSI technology. Work can be done towards applying the optimal IS filter on other orthogonal transforms. This area is very promising since many orthogonal transforms are being implemented on a single chip, such as the DCT. As a result, more optimal filter designs are needed to counter the noise existing in transform coding systems.

REFERENCES

- [Aba82] Abatzoglou, T. and O'Donnell, B., "Minimization by Coordinate Descent," *Journal of Optimization Theory and Applications*, Vol. 36, No. 2, February 1982.
- [Ahm74] Ahmed, N., Natarajan, T., and Rao, K.R., "Discrete Cosine Transform," *IEEE Transactions on Computers*, Vol. 23, pp. 90-93, January 1974.
- [Ang91] Ang, P.H., Ruetz, P.A., and Auld, D., "Video Compression Makes Big Gains," *IEEE Spectrum*, Vol. 28, No. 10, pp. 16-19, October 1991.
- [Bek60] Bekesy, G.V., *Experiments in Hearing*, New York, McGraw-Hill, 1960.
- [Bro61] Brown, W.M., "Optimal Prefiltering of Sampled Data," *IRE Transactions on Information Theory*, pp. 269-230, October 1961.
- [Buz80] Buzo, A., Gray, A.H., Gray, R.M., and Markel, J.D., "Speech Coder based upon Vector Quantization," *IEEE Transactions on Speech, Acoustics, and Signal Processing*, Vol. 28, No. 5, pp. 562-574, October 1980.
- [Cal72] Callahan, M.W. *Acoustical Signal Processing*, PhD. Dissertation, North Carolina State University, 1972.
- [Cha71] Chan, D., and Donaldson, R.W., "Optimum Pre- and Postfiltering of Sampled Signals with Application to Pulse Modulation and Data Compression Systems," *IEEE Transactions on Communication Technology*, Vol. 19, No. 2, pp. 141-156, April 1971.
- [Cha87] Chang, P.C., Gray, R.M., and May, J., "Fourier Transform Vector Quantization for Speech Coding," *IEEE Transactions on Communications*, Vol. 35, No. 10, pp. 1059-1068, October 1987.
- [Chu82] Chu, P.L., and Messerschmitt, D.G., "A Frequency Weighted Itakura-Saito Spectral Distance Measure," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 30, No. 4, pp. 545-560, August 1982.

- [Cla81] Clarke, R.J., "Relation between the Karhunen Loeve and Cosine Transforms," *IEE Proceedings, Pt.8*, Vol. 128, No. 6, pp. 359-360, November 1981.
- [Cos52] Costa, J.P., "Coding with Linear Systems," *Proceeding of the IRE*, Vol. 40, pp. 1101-1103, September 1952.
- [Cra66] Cramer, G.B., "Optimal Linear Filtering of Analog Signals in Noisy Channels," *IEEE Transactions on Audio and Electroacoustics*, Vol. 14, No. 1, pp. 3-15, March 1966.
- [Dim89] Dimolitsas, D., "Objective Speech Distortion Measures and Their Relevance to Speech Quality Assessments," *IEE proceedings*, Vol. 136, Pt. 1, No. 5, pp. 317-324, October 1989.
- [Elj87] El-Jourdi, A., and Makhoul, J., "Discrete All-Pole Modeling for Voiced Speech," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 8.9.1-8.9.4, Dallas, 1987.
- [Elj88] El-Jourdi, A., *Discrete Spectral Modeling with Application to Speech Analysis*, PhD. Dissertation, Northeastern University, 1988.
- [Elj89] El-Jourdi, A., and Makhoul, J., "Discrete Pole-Zero Modeling and Applications," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2162-2165, Glasgow, Scotland, 1989.
- [Elj91] El-Jourdi, A., and Makhoul, J., "Discrete All-Pole Modeling," *IEEE Transactions on Signal Processing*, Vol. 39, No. 2, pp. 411-423, February 1991.
- [Fla70] Flanagan, J.L., Coker, C.H., Rabiner, L.R., Schafer, R.W., and Umeda, N., "Synthesis Voices for Computers," *IEEE Spectrum*, Vol. 7, No. 10, pp. 22-45, October 1970.
- [Fla72] Flanagan, J.L., *Speech Analysis Synthesis and Perception*, 2nd Ed., Springer-Verlag, 1972.
- [Ger74] Gerber, S.E., *Introductory Hearing Science*, W.B. Saunders Company, 1974.
- [Gol78] Goldstein, M., and Dillion, W.M., *Discriminant Analysis.*, John Wiley and Sons, Inc., 1978.

- [Goo66] Goodman, L.M., and Drouilhet, P.E., "Asymptotically Optimum Pre-Emphasis and De-Emphasis Networks for Sampling and Quantizing," *Proceeding of the IEEE*, pp. 795-796, May 1966.
- [Gra76] Gray, A.H., and Markel, J.D., "Distance Measure for Speech Processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 24, No. 5, pp. 380-391, October 1976.
- [Gra80] Gray, R.M., Buzo, A., Gray, A.H., Matsuyama, Y., "Distortion Measures for Speech Processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 28, No. 4, pp. 367-376, August 1980.
- [Gra84] Gray, R.M., "Vector Quantization," *IEEE Signal Processing Magazine*, pp. 4-29, April 1984.
- [Hla92] Hlawatsch, F., and Boudreaux-Bartels, G.F., "Linear and Quadratic Time-Frequency Signal Representations," *IEEE Signal Processing Magazine*, Vol. 9, No. 2, pp. 21-67, April 1992.
- [Hol62] Hollbrook, A., and Fairbanks, G., "Diphthong Formants and Their Movements," *Journal of Speech and Hearing Research*, Vol. 5, No. 1, pp. 38-58, March 1962.
- [Ita68] Itakura, F., and Saito, S., "Analysis Synthesis Telephony Based on Maximum Likelihood Method," *International Congress on Acoustics 6th*, pp. C17-C20, Tokyo, Japan, 1968.
- [Ita70] Itakura, F., and Saito, S., "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies," *Electronic and Communication of Japan*, Vol. 53-A, pp. 36-43, 1970.
- [Ita75] Itakura, F., "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 23, No. 1, pp. 67-72, February 1975.
- [Jai89] Jain, A.K., *Fundamental of Digital Image Processing*, Prentice Hall, 1989.
- [Jay84] Jayant, N.S., and Noll, P., *Digital Coding of Waveforms*, Prentice Hall, 1984.
- [Joh88] Johnson, R.A., and Wichern, D.W., *Applied Multivariate Statistical Analysis*, 2nd Ed., Prentice-Hall, 1988.

- [Jua84] Juang, B.H., "On Using the Itakura-Saito Measures for Speech Coder Performance Evaluation," *AT&T Bell Laboratories Journal*, Vol. 63, No. 8, pp. 1477-1498, October 1984.
- [Kay88] Kay, S.M., *Modern Spectral Estimation: Theory and Application*, Prentice-Hall, 1988.
- [Kul59] Kullback, S., *Information Theory and Statistics*, John Wiley & Sons, Inc., 1959.
- [Li89] Li, J. and Krishnamurthy, A.K., "A Modified Frequency-Weighted Itakura Spectral Distortion Measure," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 10, October 1989.
- [Lin80] Linde, Y., Buzo, A., and Gray, R.M., "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, Vol. 28, No. 1, pp. 84-95, January 1980.
- [Llo57] Lloyd, S.P., "Least Squares Quantization in PCM's," *Bell Telephone Laboratories Paper*, Murray Hill, NJ, 1957.
- [Mal86] Malvar, S.H., "Optimal Pre- and Post-Filtering in Noisy Sampled-Data Systems," *RLE Technical Report No.519, Research Laboratory of Electronics*, Massachusetts Institute of Technology, August 1986.
- [Mal88] Malvar, S.H., and Staelin, D.H., "Optimal FIR Pre- and Post-Filters for Decimation and Interpolation of Random Signals," *IEEE Transactions on Communications*, Vol. 36, No. 1, January 1988.
- [Mak75] Makhoul, J., "Linear Prediction: A Tutorial Review," *Proceeding IEEE*, Vol. 63, No. 4, pp. 561-580, April 1975.
- [Mar76] Markel, J.D., and Gray, A.H., *Linear Prediction of Speech*, New York: Springer, 1976.
- [Mat87] Mathews, J.H., *Numerical Methods for Computer Science, Engineering, and Mathematics*, Prentice-Hall, 1987.
- [Noc85] Nocerino, N., Soong, F.K., Rabiner, L.R., and Klatt, D.H., "Comparitive Study of Several Distortion Measures for Speech Recognition," *Speech Communication*, Vol. 4, pp. 317-331, No. 4, December 1985.
- [Opp89] Oppenheim, A.H., and Schaffer, R.W., *Discrete-Time Signal Processing*. Prentice-Hall, 1989.

- [Orf90] Orfanidis, S.J., *Optimum Signal Processing: An Introduction*, 2nd Ed., McGraw-Hill, 1990.
- [Pap84] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, 2nd Ed., McGraw-Hill, 1984.
- [Pet52] Peterson, G.E., and Barney, H.L., "Control Methods Used in a Study of the Vowels," *Journal of Acoustic Society of America*, Vol. 24, No. 2, pp. 175-184, March 1952.
- [Pin64] Pinsker, M.S., *Information and Information Stability of Random Variables and Processes*, Holden-Day, Inc., 1964.
- [Rab75] Rabiner, L.R., and Gold, B., *Theory and Application of Digital Signal Processing*, Prentice Hall, 1975.
- [Rab78] Rabiner, L.R., and Schafer, R.W., *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
- [Rek83] Reklaitis, G.V., *Engineering Optimization: Methods and Applications*, Addison Wiley, New York, 1983.
- [Smi65] Smith, J.W., "The Joint Optimization of Transmitted Signal and Receiving Filter for Data Transmission Systems," *The Bell System Technical Journal*, pp. 2363-2392, December 1965.
- [Soo88] Soong, F.K., and Sondhi, M.M., "A Frequency-Weighted Itakura Spectral Distortion Measure and its Application to Speech Recognition in Noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 1, pp. 41-48, January 1988.
- [Tea91] Teague, K., *Real Speech Data: ECEN 5753 Digital Speech Signal Processing*, 1991.
- [Wid85] Widrow, B. and Stearns, S.D., *Adaptive Signal Processing*, Englewood Cliffs, NJ, Prentice-Hall, 1985.
- [Vai87] Vaidyanathan, P.P., "Theory and Design of M-Channel Maximally Decimated Quadrature Mirror Filters with Arbitrary M, Having the Perfect-Reconstruction Property," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 35, No. 4, pp. 476-492, April 1987.

- [Vai89] Vaidyanathan, P.P., Nguyen, T.Q., Doganata, Z., and Saramaki, T., "Improved Technique for Design of Perfect Reconstruction FIR QMF Banks with Lossless Polyphase Matrices," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 7, pp. 1042-1056, July 1989.
- [Vai90] Vaidyanathan, P.P. and Liu, V.C., "Efficient Reconstruction of Band-Limited Sequences from Nonuniformly Decimated Versions by Use of Polyphase Filter Banks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 38, No. 11, November 1990.

VITA

Chindakorn Tuchinda

Candidate for the Degree of

Doctor of Philosophy

Thesis: OPTIMAL FINITE IMPULSE RESPONSE FILTER DESIGN BASED ON MINIMIZING THE ITAKURA-SAITO DISTORTION MEASURE WITH APPLICATIONS TO DIGITAL SPEECH COMMUNICATIONS

Major Field: Electrical Engineering

Biographical:

Personal Data: Born in Bangkok, Thailand, December 8, 1964, the son of Uthai and Saisudchai Tuchinda.

Education: Graduated from Saint Gabriel's College, Bangkok, Thailand, in May 1983; received Bachelor of Engineering degree in Electrical Engineering from Chulalongkorn University in May, 1986; received Master of Science degree from Oklahoma State University in May 1989; complete the requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1992.

Professional Experience: Teaching and Research Assistant, Department of Electrical and Computer Engineering, Oklahoma State University, August 1988, to December, 1992; Member of Tau Beta Pi and Eta Kappa Nu.