Low Distortion Speech Enhancement in the DCT Domain using Optimal Estimate of the *a priori* SNR

by

Lutfa Akter

A thesis submitted to the Department of Electrical and Electronic Engineering
of
Bangladesh University of Engineering and Technology
in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONIC ENGINEERING

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
BANGLADESH UNIVERSITY OF ENGINEERING AND TECHNOLOGY
February 2004

The thesis entitled **"Low Distortion Speech Enhancement in the DCT Domain using Optimal Estimate of the *a priori* SNR"** submitted by Lutfa Akter Roll No.: 040206217P, Session: April, 2002 has been accepted as satisfactory in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronic Engineering** on February 26, 2004.

BOARD OF EXAMINERS

1. _____ **Chairman**
(Dr. Md. Kamrul Hasan)
Associate Professor
Department of Electrical and
Electronic Engineering, BUET,
Dhaka-1000, Bangladesh

2. _____ **Member**
26/02/04
(Dr. Newaz Muhammad Syfur Rahim)
Assistant Professor
Department of Electrical and
Electronic Engineering, BUET,
Dhaka-1000, Bangladesh

3. _____ **Member**
(Ex-officio)
(Dr. Mohammad Ali Choudhury)
Professor and Head
Department of Electrical and
Electronic Engineering, BUET,
Dhaka-1000, Bangladesh

4. _____ **Member**
(External)
(Dr. Farruk Ahmed)
Professor
Department of Computer Science
and Engineering,
North South University,
Dhaka-1213, Bangladesh

Declaration

It is hereby declared that this thesis or any part of it has not been submitted elsewhere for the award of any degree or diploma.

Signature of the candidate

Lutfa 21/3/2004

(Lutfa Akter)

Dedication

To my beloved parents.

Acknowledgements

I would like to express my sincere gratitude and profound indebtedness to Dr. Md. Kamrul Hasan, Associate Professor, Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka, for his constant guidance, constructive suggestions, commendable support and endless patience throughout the progress of this work, without which the work could not have been completed.

I am also thanking my friend Zahidur Rahim Chowdhury for encouraging me to choose my supervisor and the subject.

Contents

List of Tables

List of Figures

Abstract

The spectral subtraction based algorithms are commonly used for single channel speech enhancement because of their elegant performance in denoising with low computational load. They, however, suffer from a serious drawback in that the enhanced speech is accompanied by unpleasant *musical noise* artifact, which is characterized by tones with random frequencies. It is known that the key point behind the reduction of *musical noise* by the minimum-mean-squared-error (MMSE) estimator is the use of *a priori* SNR. The "decision-directed" approach widely used for its estimation requires an averaging parameter. Conventionally, a constant value is chosen by most researchers. The main objective of this work is the development of a self-adaptive smoothing parameter in the MMSE sense to estimate the *a priori* SNR in the DCT domain which can account for the abrupt changes in the speech spectral amplitudes. The performance improvement using the proposed self-adaptive smoothing parameter in the commonly used spectral subtraction algorithms for denoising speech corrupted by background noise is noteworthy.

The conventional Wiener filtering shows better denoising performance in terms of overall and average segmental SNRs with the cost paid in Itakura-Saito (IS) measure as compared to the spectral subtraction based methods. In this work, a generalized Wiener filter is proposed to improve the IS measure without sacrificing enhanced speech quality in terms of SNR by introducing a new term in the gain function. A comparative study with the spectral subtraction algorithms and the conventional Wiener filter confirms the superiority of the proposed generalized Wiener filter.

Chapter 1

Introduction

1.1 Speech Enhancement: Background

Speech enhancement is the term used to describe algorithms or devices whose purpose is to improve some perceptual aspects of speech for the human listener or to improve the speech signal so that it may be better exploited by other speech processing algorithms. Development and widespread deployment of digital communication systems during the last twenty years have brought increased attention to the role of speech enhancement in speech processing problems [1]-[6]. Speech enhancement algorithms have been applied to problems as diverse as correction of reverberation, pitch modification, rate modification, reconstruction of lost speech packets in digital networks, correction of so-called "hyperbaric" speech produced by deep-sea divers breathing a helium-oxygen mixture and correction of speech that has been distorted due to pathological problems of the speaker. However, noise reduction is probably the most important and most frequently encountered speech enhancement problem.

Speech enhancement attempts to improve the performance of voice communication systems when their input or output signal is corrupted by noise. The improvement is in the sense of minimizing the effects of noise on the performance of these systems. The need for enhancing speech signals arises in many situations in which the speech either originates from some noisy location or is affected by the noise over the channel or at the receiving end. Both digital and analog channels are possible, and communication can be either between people or with a machine. Hence speech enhancement is the problem of enhancing a given sample function of noisy speech signal, as well as the problem of enhancing the performance of

speech coding and recognition systems whose input signal is noisy [1]-[40]. Examples of important applications of speech enhancement include improving the performance of 1) cellular radio telephone systems, which usually suffer from background noise in the car as well as from channel noise; 2) pay phones located in noisy environments (e.g. airports); 3) air-ground communication systems in which the cockpit noise corrupts the pilot's speech; 4) teleconferencing systems where noise sources in one location may be broadcast to all other locations; 5) long distance communication over noisy radio channels; 6) paging systems located in noisy environments (e.g. airports, machine rooms); 7) ground-air communication in which the cockpit noise corrupts the received massages; and 8) suboptimal speech quantization systems.

In the cellular radio telephone example, the original speech is corrupted by the noise generated by the engine, fan, traffic and wind as well as by the channel noise [7], [8]. The signals delivered by cellular systems may therefore be noisy with impaired quality and intelligibility. If the cellular system encodes the signal prior to its transmission, then further degradation in its performance results, since speech coders rely on some model for the clean signal and normally that model is not suitable for the noisy signal. Similarly, if the cellular system is equipped with a speech recognition system which is used for automatic dialing, then the recognition accuracy of such system deteriorates in the presence of noise, since the noisy input is unlikely to obey the statistical model for the clean signal used by the recognizer. Similar problems are encountered with pay phone communication, air-ground communication, and teleconferencing systems. In the air-ground communication examples, however, the messages of low quality and intelligibility delivered to the air traffic controllers may have disastrous effects. The situation in long distance communication, paging systems, and ground-air communication is somewhat simpler, since the noise is added to the speech at the channel and at the receiving end, respectively, rather than that at the source location. Hence, the clean signal can be "immunized" prior to being affected by the noise [9]-[11]. In suboptimal quantization of speech signals, the quantized signal is considered a noisy version of the clean signal [12]-[13]. Hence, enhancement can be applied to reduce the quantization noise, provided that quantization was not optimally performed.

The forgoing discussion demonstrates that speech enhancement has three major goals: 1) to improve perceptual aspects (e.g., quality, intelligibility) of a given sample function of degraded speech signal; 2) to increase robustness of speech coders to input noise; 3) to increase robustness of speech recognition systems to input noise.

The quality of speech signal is a subjective measure which reflects on the way the signal is perceived by listeners. It can be expressed in terms of how pleasant the signal sounds or how much effort is required on behalf of the listeners in order to understand the message. Intelligibility, on the other hand, is an objective measure of the amount of information which can be extracted by listeners from the given signal, whether the signal is clean or noisy. A given signal may be of the high quality and low intelligibility, and vice versa. Hence, the two measures are independent of each other. Both the quality and the intelligibilty of a set of given signals are evaluated based on tests performed on human listeners. Since no mathematical quantification of these measures, in terms of closed-form perceptually meaningful distortion measures, is known, algorithms for goals 1 and 2 above are difficult to design and evaluate. Goal 3 is significantly simpler since the problem is that of decoding the signal into a finite number of classes and the ultimate goal can be easily formulated in mathematical terms. Usually the problem is that of designing decoders which minimize the probability of recognition error.

The speech enhancement problem consists of a family of subproblems characterized by the type of noise source, the way the noise interacts with the clean signal, the number of voice channels, or microphone outputs, available for enhancement, and the nature of speech communication systems. The noise, or the interfering signals, may, for example be due to competitive speakers, background sounds (music, fans, machines, door slamming, wind, traffic etc.), room reverberation, or random channel noise. The noise may accompany the original signal at the source location, over communication channels, or at the receiving end. It may affect the original signal in an additive, multiplicative, or convolutional manner. Furthermore, the noise may be statistically dependent or independent of the clean signal. The number of voice channels available for enhancementis an important factor in designing speech enhancement systems. In general, the

larger the number of microphones, the easier the speech enhancement task. The communication system for which speech enhancement is designed can simply be a recording which has to be displayed to audience, a man-machine communication system (speech recognizer), a digital communication system, etc.

Speech enhancement based on spectral decomposing and filtering [14]-[22] remains a common and effective approach for enhancing speech degraded by acoustic additive noise when only the noisy speech is available. This general class is based on variations of optimum filters and encompasses such methods as spectral subtraction, Wiener filtering and various maximum likelihood (ML) estimation schemes. A common set of requirements in this class include: 1) An appropriate suppression rule based on an optimality criteria [15], [16] and usually function of the SNR (signal to noise ratio) and other speech and noise statistics. 2) An estimation of the speech and noise power spectral densities, or their respective autocorrelation. 3) A quantification of the probability of speech presence to further attenuate non-speech bands [17]. 4) A method for reducing residual noise by appropriately smoothing the estimated quantities [15] and/or exploiting the psychoacoustic properties of human hearing.

The choice of suppression rules is governed by many factors, such as computational efficiency, optimality criteria, and the exploiting of human hearing properties. In the reported literature, the range includes heuristic rules (e.g., [16]) as well as formally derived ones. The ML estimation approaches in [15], [18] attempt to better exploit the statistical properties of the DFT (discrete Fourier transform) of the noisy speech. These methods assume a statistical model for the DFT coefficients of noisy speech and derive optimum estimators of the magnitude spectrum based on that model.

An important contribution in this area is the smoothing approach proposed in [15] whereby the variation in SNR between successive frames is reduced by averaging the locally computed SNR (SNR_{post}) with the SNR estimated in the previous frame after the filtering operation (SNR_{est}). The method results in a significant reduction of the "musical noise" artifacts, as shown in [14].

Another speech enhancement approach is the signal subspace (SS) method [23], [24]. The key idea is to decompose the vector space of the noisy signal into a signal-plus-noise subspace and a noise subspace under the assumption that the

additive noise is white. The enhancement is performed by removing the noise subspace and estimating the clean speech from the remaining signal-plus-noise subspace. Hidden Markov Model (HMM) based speech enhancement approaches [25], [26] have also drawn much attention in recent years.

Methods for speech enhancement have also been developed based on extraction of parameters from noisy speech, and synthesizing speech from these parameters [27]. All-pole modelling of degraded speech is one such method [28]. In all-pole modelling, if wrong peaks are extracted, then these peaks may get enhanced. Temporal sequence of these peaks also produces discontinuities in the contours of the spectral peaks when compared with the smooth contours in natural speech. Methods for speech enhancement have also been suggested based on the periodicity due to pitch [29]. Noise samples in successive glottal cycles are uncorrelated. On the other hand, the characteristics of the vocal tract system are highly correlated due to slow movement of the articular. These methods for enhancement of speech depend critically on the estimation of pitch from the noisy speech signal.

Many speech enhancement algorithms make use of DFT to make it easier to remove noise embedded in the noisy speech signal [1]-[22]. Recently, Discrete Cosine transform (DCT) and Wavelet transform have been widely used as analysis tools in the field of speech enhancement [30]-[36]. DCT is widely used because of its excellent energy compaction properties.

1.2 Objective of This Research

This main objective of this thesis work is the development of a self adaptive smoothing parameter in the MMSE sense to estimate the *a priori* SNR in the DCT domain which can account for the abrupt changes in the speech spectral amplitude in the spectral subtraction approach. For single channel speech enhancement, the spectral subtraction based algorithms are commonly used because of their low computational load. However, the spectral subtraction algorithms have a serious drawback in that the enhanced speech is accompanied by unpleasant musical noise artifact, which is characterized by tones with random frequencies [1]. Apart from being extremely annoying to the listeners, the musical noise also hampers the performance of the speech-coding algorithms to a great extent [42].

It has been shown in [17] that the key point behind the reduction of *musical noise* by the minimum-mean-squared-error (MMSE) estimator [15] is the use of *a priori* SNR. Several methods such as spectral subtraction based algorithms, Wiener filtering require the knowledge of the *a priori* SNR and the estimation of *a priori* SNR using the "decision-directed" approach requires an averaging parameter [17]. A low value of the averaging parameter is suitable for rapidly changing speech regions, while a high value is suitable for near stationary speech frames [43]. Conventionally, a constant value is used for the averaging parameter [30].

A self adaptive optimum smoothing parameter to estimate the *a priori* SNR in the DCT domain is derived in Chapter 3. The performance of the proposed self adaptive smoothing parameter on the commonly used spectral subtraction algorithms is evaluated on speech corrupted by background white Gaussian noise and color noise (e.g., babble noise). It is also expected that incorporation of improved estimate of the *a priori* SNR in the variants of the Wiener filtering algorithm will significantly improve their performances both in terms of quality and intelligibility.

Though the Wiener filtering shows better performance in terms of overall output SNR and average segmental SNR (AvgSegSNR), its IS measure is the worst as compared to the spectral subtraction based methods. To improve the IS measure without sacrificing SNR, a generalized Wiener filter is proposed by relaxing a basic assumption in Chapter 4. A comparative study with the spectral subtraction algorithms and the Wiener filtering is provided to demonstrate the effectiveness of implementing the proposed generalized Wiener filter.

In this work, we have computed noise spectral components from noisy speech according to [41]. Here all computations are done in the DCT domain. The transform coefficients are first divided into a number of blocks consisting of convenient number of consecutive coefficients of the transformed signal.

1.3 Organization of the Thesis

This thesis consists of five chapters. Chapter 1 gives a brief description of necessity of speech enhancement techniques, names of existing methods and the main objectives of this research work.

In Chapter 2, a brief review of previous differents peech enhancement techniques such as spectral subtraction rules, Wiener filtering and dual gain Wiener filtering are presented.

In Chapter 3, the drawback of the traditional spectral subtraction based methods is discussed. To improve their performances, more accurate method for *a priori* SNR estimation is proposed. Simulation results are also presented to evaluate the effectiveness of the proposed scheme.

In Chapter 4, a generalized gain function for the Wiener filter is proposed. The performance of the proposed generalized Wiener filter is evaluated and compared with that of the conventional Wiener filter.

The thesis concludes by presenting an overall discussion on the work and pointing out some unsolved problems for future work in Chapter 5.

Chapter 2

Review of Speech Enhancement Techniques

2.1 Introduction

Speech enhancement plays a key role in designing robust automatic speech and speaker recognition systems. As the presence of noise deteriorates the performance of the recognition systems and also shows an adverse effect on the perceived quality and intelligibility of speech at the receiving end, several approaches for speech enhancement in additive noise have been proposed. Speech enhancement based on spectral decomposition and variations of optimum filters cover the methods such as spectral subtraction, Wiener filtering and various maximum likelihood (ML) estimation schemes. The major breakthrough in speech enhancement technique are described in the following sections.

2.2 Enhancement Techniques Based on Spectral Amplitude Estimation

In general, in enhancement of a signal degraded by additive noise, it is significantly easier to estimate the spectral amplitude associated with the original signal than it is to estimate both amplitude and phase. It is principally the spectral amplitude rather than phase that is important for speech intelligibility and quality. There are a variety of speech enhancement techniques that capitalize on this aspect of speech perception by focusing on enhancing only the spectral amplitude. The techniques to be discussed can be broadly classified into two groups. First, the spectral amplitude is estimated in the frequency domain, using the spectrum

of the degraded speech. Each short-time segment of the enhanced speech waveform in the time domain is then obtained by inverse transforming this spectral amplitude estimate combined with the phase of the degraded speech. In the second class, the degraded speech is first used to obtain a filter which is then applied to the degraded speech. Since these procedures lead to zero-phase filters, it is again only the spectral amplitude that is enhanced, with the phase of the filter being identical to that of the degraded speech.

2.2.1 Spectral amplitude estimation based on spectral subtraction

A classical noise reduction approach for speech enhancement and robust recognition is the spectral subtraction method that was first proposed by Boll [1]. The basic idea is restore the magnitude spectrum or power spectrum of a signal observed in additive noise through subtraction of an estimate of the average noise spectrum from the noisy signal spectrum. Assuming that the noise is a stationary or a slowly varying process, the noise spectrum is estimated or updated during the periods when the speech signal is absent. The estimation is performed on a frame-by-frame basis, where each frame consists 20-40 ms of speech samples. The sample spectrum of the noisy signal is usually employed in the spectral subtraction approach, thus resulting in an estimate of the sample spectrum of the clean signal. The square root of the estimate of the sample spectrum is considered an estimate of the magnitude spectrum of the speech signal. An spectral estimate of the signal is obtained by multiplying the estimate of magnitude spectrum with sign of the noisy signal in the DCT domain.

Let $x(t)$ denotes the clean signal and $d(t)$ denotes the additive noise sequence in the time domain, and it is assumed that $x(t)$ and $d(t)$ are uncorrelated. The noisy signal in the time domain $y(t)$ is given by

$$y(t) = x(t) + d(t) \tag{2.1}$$

The discrete Cosine transform (DCT) of the noisy signal $y(t)$ [38] is denoted by $Y_{n,k}$, $0 \leq k \leq N-1$, where n and k denote frame and frequency index, respectively. The sample spectrum of $y(t)$ is given by $Y_{n,k}^2$. Let $X_{n,k}$ and $D_{n,k}$ denote the clean speech and noise signal spectral component in the DCT domain,

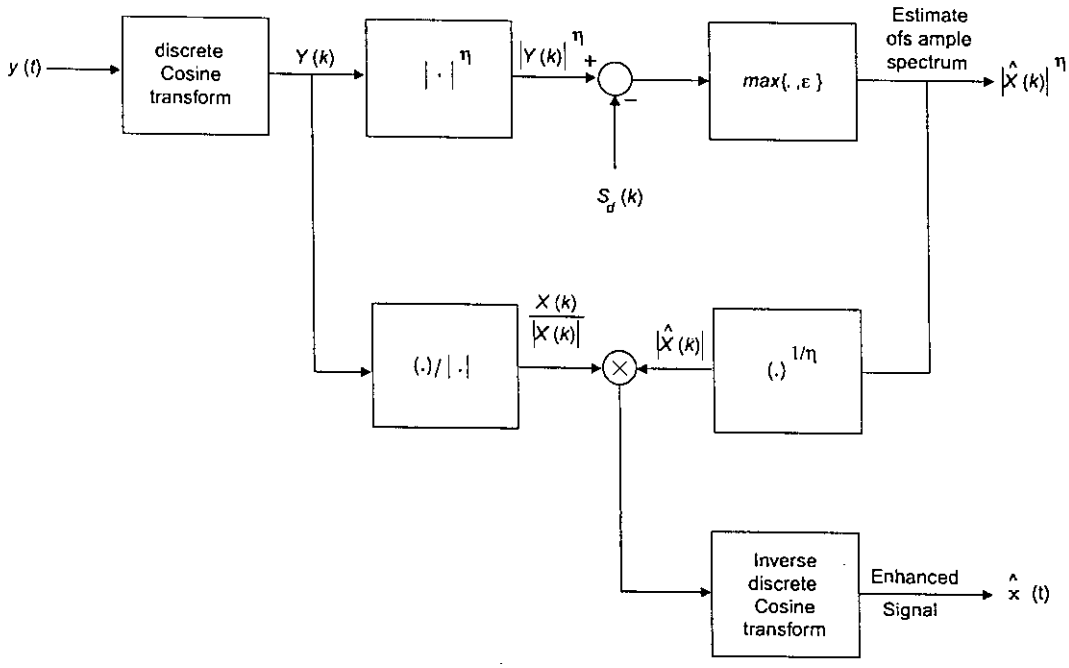


Fig. 2.1: The spectral subtraction approach.

respectively, then the DCT domain representation of Eq. (2.1) is given by

$$Y_{n,k} = X_{n,k} + D_{n,k} \tag{2.2}$$

The forward DCT of the noisy signal $\{y(t), 0 \leq t \leq N-1\}$ is given by [30]

$$Y_k = \alpha_k \sum_{t=0}^{N-1} y(t) \cos\left[\frac{\pi(2t+1)k}{2N}\right], 0 \leq k \leq N-1 \tag{2.3}$$

where

$$\alpha_k = \begin{cases} \sqrt{\frac{1}{N}}, k = 0 \\ \sqrt{\frac{2}{N}}, 1 \leq k \leq N-1 \end{cases} \tag{2.4}$$

The reconstructed signal, $\widehat{x}(t)$, can be obtained using the following inverse Cosine transformation (IDCT) [30]

$$\widehat{x}(t) = \sum_{k=0}^{N-1} \alpha_k \widehat{X}_k \cos\left[\frac{\pi(2t+1)k}{2N}\right], 0 \leq t \leq N-1 \tag{2.5}$$

where $\widehat{X}_k$ denotes the denoised DCT coefficients. A block diagram of the spectral subtraction approach is shown in Fig. 2.1, η can be of 1 or 2, $\eta = 2$ provides best result; $S_d(k)$ is the noise estimate.

With the exact noise spectrum $D_{n,k}$ the "ideal" spectral power subtraction takes the form [20], [30], [41], i.e.,

$$Y_{n,k}^2 = X_{n,k}^2 + D_{n,k}^2 \tag{2.6}$$

Rearranging Eq. (2.6)

$$\begin{aligned} X_{n,k}^2 &= Y_{n,k}^2 - D_{n,k}^2 \\ &= Y_{n,k}^2 \left\{ 1 - \frac{D_{n,k}^2}{Y_{n,k}^2} \right\} \end{aligned} \tag{2.7}$$

Thus in the DCT domain, power of the clean speech spectral is estimated as

$$X_{n,k}^2 = Y_{n,k}^2 \left\{ 1 - \frac{E\{D_{n,k}^2\}}{Y_{n,k}^2} \right\} \tag{2.8}$$

and the clean speech spectral component itself is obtained as

$$\begin{aligned} \widehat{X}_{n,k} &= \sqrt{Y_{n,k}^2 \left\{ 1 - \frac{E\{D_{n,k}^2\}}{Y_{n,k}^2} \right\}} \\ &= Y_{n,k} \sqrt{\left\{ 1 - \frac{E\{D_{n,k}^2\}}{Y_{n,k}^2} \right\}} \end{aligned} \tag{2.9}$$

where $E\{\cdot\}$ is the expectation operator. Then an estimate of the clean speech spectral component in the DCT domain in direct spectral power subtraction approach can be written as

$$\widehat{X}_{n,k} = G_{n,k}.Y_{n,k} \tag{2.10}$$

where $G_{n,k}$ is called the gain function, given by

$$G_{n,k} = \sqrt{max\left\{0, \left(1 - \frac{E\{D_{n,k}^2\}}{Y_{n,k}^2}\right)\right\}} \tag{2.11}$$

provided that the difference of spectral estimates of the noisy signal and the noise process is nonnegative. If this difference becomes negative, then it is usually replaced by an arbitrary small nonnegative number, ϵ. The power spectral density of the noise is normally estimated from portions of the noisy signal during when speech is absent and only noise is present. The spectral subtraction based signal estimator affects the magnitude spectrum of the noisy signal in each frame while it keeps the phase of that signal intact. From a perceptual point of view this is a desirable property, since the short-time magnitude spectrum of the clean signal is considerably more important than its short-time phase spectrum [2], [3], [39] and optimal estimation of the short-time magnitude and phase spectrum of the clean signal cannot be simultaneously performed [15], [25]. Many variations on the basic spectral subtraction approach have been proposed [1], [2], [18], [20], [40].

Scalart and Filho proposed estimation of the *a priori* SNR [17]. The local *a posteriori* SNR and *a priori* SNR are defined as follows:

$$\mathrm{SNR}_{post}(n,k) = \gamma_{n,k} = \frac{Y_{n,k}^2}{\sigma_d^2(n,k)} \tag{2.12}$$

$$\mathrm{SNR}_{prior}(n,k) = \xi_{n,k} = \frac{E\{X_{n,k}^2\}}{\sigma_d^2(n,k)} \tag{2.13}$$

where $\sigma_d^2(n,k) = E\{D_{n,k}^2\}$. An estimation of $\xi_{n,k}$ is made according to the "decision-directed" approach by Ephraim and Malah [15]

$$\widehat{\xi}_{n,k} = \alpha \frac{\widehat{X}_{n-1,k}^2}{\widehat{\sigma}_d^2(n-1,k)} + (1-\alpha)P[\gamma_{n,k} - 1] \tag{2.14}$$

where $\widehat{X}_{n-1,k}$ is the estimate of the kth speech spectral component and $\widehat{\sigma}_d^2(n-1,k)$ is the estimate of variance of the kth noise spectral component in the $(n-1)$th analysis frame, the operator $P[\cdot]$ denotes half wave rectification, and α is an averaging parameter. Considering the maximum likelihood estimate of the *a priori* SNR, we have $\xi_{n,k} = E\{\gamma_{n,k} - 1\}$ [15]. Therefore the gain function of Eq. (2.10) takes the form

$$G_{n,k}^{PE} = \sqrt{\frac{\widehat{\xi}_{n,k}}{(1+\widehat{\xi}_{n,k})}} \tag{2.15}$$

The method of estimating the clean speech spectral amplitude using the above gain function will be called spectral power estimation (PE) method in the following sections and chapters. Hence the final form of the above subtraction rule is

$$\widehat{X}_{n,k} = G_{n,k}^{PE}.Y_{n,k} \tag{2.16}$$

Scalart and Filho [17] also proposed to consider the possibility of the probability of speech presence in the spectral subtraction rules. Several authors incorporated the further attenuation based on the probability of speech presence [44], [45] in the FFT (fast Fourier transform) domain. Let $q_{n,k}$ is the probability of speech absence in the k-th spectral component and n-th frame. The analysis relies on a two-state model of a speech event such that the noisy signal under consideration may or may not have speech present in it. This is illustrated by the following two hypotheses:

H_0: The speech is absent: $Y_k = D_k$

H_1: The speech is present: $Y_k = X_k + D_k$

The probability that the speech is in state H_1 may be obtained according to Bayes rule as

$$\begin{aligned} p(H_1|Y_{n,k}) &= \frac{p(Y_{n,k}|H_1)p(H_1)}{p(Y_{n,k}|H_1)p(H_1)+p(Y_{n,k}|H_0)p(H_0)} \\ &= \frac{p(H_1)}{p(H_1)+p(H_0)\frac{p(Y_{n,k}|H_0)}{p(Y_{n,k}|H_1)}} \end{aligned} \tag{2.17}$$

given $Y_{n,k}$. Since $X_{n,k}$ and $D_{n,k}$ are zero-mean real Gaussian random process in the DCT domain, the probability density function of $Y_{n,k}$ follows the Gaussian distribution, i.e.,

$$p(Y_{n,k}|H_1) = \frac{1}{\sqrt{2\pi E\{Y_{n,k}^2\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2\}}\right) \tag{2.18}$$

$$\begin{aligned} p(Y_{n,k}|H_0) &= \frac{1}{\sqrt{2\pi E\{Y_{n,k}^2\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2\}}\right) \\ &= \frac{1}{\sqrt{2\pi E\{D_{n,k}^2\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{D_{n,k}^2\}}\right) \end{aligned} \tag{2.19}$$

$p(Y_{n,k}|H_0)/p(Y_{n,k}|H_1)$ is obtained as

$$\begin{aligned} \frac{p(Y_{n,k}|H_0)}{p(Y_{n,k}|H_1)} &= \sqrt{\frac{E\{Y_{n,k}^2\}}{E\{D_{n,k}^2\}}} \exp\left(-\frac{Y_{n,k}^2}{2E\{D_{n,k}^2\}}+\frac{Y_{n,k}^2}{2E\{Y_{n,k}^2\}}\right) \\ &= \sqrt{\frac{E\{Y_{n,k}^2\}}{E\{D_{n,k}^2\}}} \exp\left(-\frac{\gamma_{n,k}}{2}+\frac{Y_{n,k}^2}{2\left(E\{(X_{n,k}^2+D_{n,k}^2)\}\right)}\right) \\ &= \sqrt{\gamma_{n,k}} \exp\left(-\frac{\gamma_{n,k}}{2}+\frac{Y_{n,k}^2}{2\left(E\{X_{n,k}^2\}+E\{D_{n,k}^2\}\right)}\right) \\ &= \sqrt{\gamma_{n,k}} \exp\left(-\frac{\gamma_{n,k}}{2}+\frac{\frac{Y_{n,k}^2}{E\{D_{n,k}^2\}}}{2\left(\frac{E\{X_{n,k}^2\}}{E\{D_{n,k}^2\}}+\frac{E\{D_{n,k}^2\}}{E\{D_{n,k}^2\}}\right)}\right) \\ &= \sqrt{\gamma_{n,k}} \exp\left(-\frac{\gamma_{n,k}}{2}+\frac{\gamma_{n,k}}{2(\xi_{n,k}+1)}\right) \\ &= \sqrt{\gamma_{n,k}} \exp\left(\frac{\gamma_{n,k}(-\xi_{n,k}-1+1)}{2(\xi_{n,k}+1)}\right) \\ &= \sqrt{\gamma_{n,k}} \exp\left(-\frac{\gamma_{n,k}\xi_{n,k}}{2(\xi_{n,k}+1)}\right) \end{aligned} \tag{2.20}$$

Replacing $p(H_0) = q_{n,k}$, and $p(H_1) = 1 - q_{n,k}$, and defining $p(H_1|Y_{n,k}) = G_{n,k}^{pr}$, we get in the DCT domain

$$G_{n,k}^{pr} = \frac{1-q_{n,k}}{1-q_{n,k}+q_{n,k}\sqrt{\gamma_{n,k}}\exp\left(-\frac{\gamma_{n,k}\xi_{n,k}}{2(\xi_{n,k}+1)}\right)}$$

$$= \frac{1-q_{n,k}}{1-q_{n,k}+q_{n,k}\sqrt{\gamma_{n,k}}\exp(-\nu_{n,k})} \tag{2.21}$$

Thus an additional attenuation based on the uncertainty of the speech presence in the DCT domain is

$$G_{n,k}^{pr} = \frac{1-q_{n,k}}{1-q_{n,k}+q_{n,k}\sqrt{(1+\xi_{n,k})}\exp(-\nu_{n,k})} \tag{2.22}$$

where $\nu_{n,k} = \gamma_{n,k}\xi_{n,k}/2(1+\xi_{n,k})$. $\xi_{n,k}$ is calculated by using Eq. (2.14). Thus the effective gain in this method can be expressed as

$$G_{n,k}^{MPE} = \frac{1-q_{n,k}}{1-q_{n,k}+q_{n,k}\sqrt{(1+\xi_{n,k})}\exp(-\nu_{n,k})} G_{n,k}^{PE} \tag{2.23}$$

The corresponding subtraction rule takes the form

$$\widehat{X}_{n,k} = G_{n,k}^{MPE}.Y_{n,k} \tag{2.24}$$

This method will be called modified spectral power estimation (MPE) in the following sections and chapters.

In the recent years efforts were directed to find an optimum parametric estimator [20] by using statistical distribution of speech and noise signals and optimizing in the MMSE sense. The authors [20] estimated the clean speech spectral power as

$$\begin{aligned} \widehat{X}_{n,k}^2 &= a_{n,k}Y_{n,k}^2 - a_{n,k}E\{D_{n,k}^2\} \\ &= a_{n,k}\left\{Y_{n,k}^2 - E\{D_{n,k}^2\}\right\} \end{aligned} \tag{2.25}$$

By using Eqs. (2.8)-(2.15), an estimate of clean speech spectral component is obtained as

$$\widehat{X}_{n,k} = \sqrt{a_{n,k}}\sqrt{\frac{\widehat{\xi}_{n,k}}{(1+\widehat{\xi}_{n,k})}}.Y_{n,k} \tag{2.26}$$

"a" is evaluated in the minimum mean-square (MMSE) sense by minimizing the cost function

$$\delta = E\left\{(X_{n,k}^2 - \widehat{X}_{n,k}^2)^2\right\} \tag{2.27}$$

Substituting Eqs. (2.25) and (2.7) consecutively into Eq. (2.27)

$$\delta = E\left\{\left[X_{n,k}^2 - a_{n,k}Y_{n,k}^2 + a_{n,k}E\{D_{n,k}^2\}\right]^2\right\}$$

$$
\begin{aligned}
&= E\left\{\left[X_{n,k}^2 - a_{n,k}(X_{n,k}^2 + D_{n,k}^2) + a_{n,k}E\{D_{n,k}^2\}\right]^2\right\} \\
&= E\left\{\left[(1-a_{n,k})X_{n,k}^2 - a_{n,k}D_{n,k}^2 + a_{n,k}E\{D_{n,k}^2\}\right]^2\right\} \\
&= E\left\{\left[(1-a_{n,k})^2X_{n,k}^4 + a_{n,k}^2D_{n,k}^4 + a_{n,k}^2E\{D_{n,k}^2\}^2\right.\right. \\
&\quad -2a_{n,k}(1-a_{n,k})X_{n,k}^2D_{n,k}^2 - 2a_{n,k}^2D_{n,k}^2E\{D_{n,k}^2\} \\
&\quad \left.\left.+2a_{n,k}(1-a_{n,k})X_{n,k}^2E\{D_{n,k}^2\}\right]\right\} \\
&= (1-a_{n,k})^2E\{X_{n,k}^4\} + a_{n,k}^2E\{D_{n,k}^4\} + a_{n,k}^2E\{D_{n,k}^2\}^2 \\
&\quad -2a_{n,k}(1-a_{n,k})E\{X_{n,k}^2\}E\{D_{n,k}^2\} \\
&\quad -2a_{n,k}^2E\{D_{n,k}^2\}E\{D_{n,k}^2\} \\
&\quad +2a_{n,k}(1-a_{n,k})E\{X_{n,k}^2\}E\{D_{n,k}^2\} \\
&= (1-a_{n,k})^2E\{X_{n,k}^4\} + a_{n,k}^2E\{D_{n,k}^4\} + a_{n,k}^2E\{D_{n,k}^2\}^2 \\
&\quad -2a_{n,k}(1-a_{n,k})E\{X_{n,k}^2\}E\{D_{n,k}^2\} - 2a_{n,k}^2E\{D_{n,k}^2\}^2 \\
&\quad +2a_{n,k}(1-a_{n,k})E\{X_{n,k}^2\}E\{D_{n,k}^2\} \\
&= (1-a_{n,k})^2E\{X_{n,k}^4\} + a_{n,k}^2E\{D_{n,k}^4\} - a_{n,k}^2E\{D_{n,k}^2\}^2 \qquad (2.28)
\end{aligned}
$$

Differentiating δ with respect to $a_{n,k}$ gives

$$
\begin{aligned}
\frac{\partial\delta}{\partial a_{n,k}} = &\quad 2(1-a_{n,k})(-1)E\{X_{n,k}^4\} + 2a_{n,k}E\{D_{n,k}^4\} - 2a_{n,k}E\{D_{n,k}^2\}^2 \\
&= 2a_{n,k}E\{X_{n,k}^4\} + 2a_{n,k}E\{D_{n,k}^4\} - 2a_{n,k}E\{D_{n,k}^2\}^2 \\
&\quad -2E\{X_{n,k}^4\} \qquad (2.29)
\end{aligned}
$$

Equating $\partial\delta/\partial a_{n,k}$ to zero yields an optimum expression for $a_{n,k}$

$$
a_{n,k} = \frac{E\{X_{n,k}^4\}}{E\{X_{n,k}^4\} + E\{D_{n,k}^4\} - E\{D_{n,k}^2\}^2} \qquad (2.30)
$$

Since $X_{n,k}$ and $D_{n,k}$ in the DCT domain are zero-mean real Gaussian random process, the probability density function of $X_{n,k}$ follows the Gaussian distribution, i.e.,

$$
p(X_{n,k}) = \frac{1}{\sqrt{2\pi E\{X_{n,k}^2\}}}\exp\left(-\frac{X_{n,k}^2}{2E\{X_{n,k}^2\}}\right) \qquad (2.31)
$$

Fourth moment of $X_{n,k}$, i.e. $E\{X_{n,k}^4\}$, is then given by

$$
\begin{aligned}
E\{X_{n,k}^4\} &= \int_{-\infty}^{\infty} X_{n,k}^4 p(X_{n,k})dX_{n,k} \\
&= \frac{2}{\sqrt{2\pi E\{X_{n,k}^2\}}}\int_0^{\infty} X_{n,k}^4\exp\left(-\frac{X_{n,k}^2}{2E\{X_{n,k}^2\}}\right)dX_{n,k} \qquad (2.32)
\end{aligned}
$$

Now, using the formula

$$\int_0^\infty x^m \exp(-ax^2)dx = \frac{\Gamma(\frac{m+1}{2})}{2a^{\frac{m+1}{2}}},\ a>0,\ m>-1 \tag{2.33}$$

Thus $E\{X^4\}$ is obtained as (dropping subscript (n,k) for notational simplicity)

$$\begin{aligned} E\{X^4\} &= \frac{2}{\sqrt{2\pi E\{X^2\}}} \frac{\Gamma(\frac{4+1}{2})}{2(1/2E\{X^2\})^{\frac{4+1}{2}}} \\ &= \frac{2}{\sqrt{2\pi E\{X^2\}}} \frac{\Gamma(\frac{5}{2})}{2(1/2E\{X^2\})^{\frac{5}{2}}} \\ &= \frac{1}{\sqrt{2\pi E\{X^2\}}} \frac{\frac{3}{4}\sqrt{\pi}}{2(1/2E\{X^2\})^{\frac{5}{2}}} \\ &= \frac{4\sqrt{2}E\{X^2\}^2}{\sqrt{2\pi}} \frac{\frac{3}{4}\sqrt{\pi}}{1} \\ &= 3E\{X^2\}^2 \end{aligned} \tag{2.34}$$

as $\Gamma(n+1) = n\Gamma(n)$. We obtain

$$E\{X_{n,k}^4\} = 3E\{X_{n,k}^2\}^2 \tag{2.35}$$

Similarly,

$$E\{D_{n,k}^4\} = 3E\{D_{n,k}^2\}^2 \tag{2.36}$$

Substituting Eqs. (2.35) and (2.36) into Eq. (2.30), we get

$$\begin{aligned} a_{n,k} &= \frac{3E\{X_{n,k}^2\}}{3E\{X_{n,k}^2\} + 3E\{D_{n,k}^2\} - E\{D_{n,k}^2\}^2} \\ &= \frac{3E\{X_{n,k}^2\}}{3E\{X_{n,k}^2\} + 2E\{D_{n,k}^2\}} \\ &= \frac{\xi_{n,k}^2}{\xi_{n,k}^2 + 0.667} \end{aligned} \tag{2.37}$$

Finally, the corresponding subtraction rule (Eq. (2.26)) can be expressed as

$$\begin{aligned} \widehat{X}_{n,k} &= \sqrt{\frac{\xi_{n,k}^2}{0.667 + \xi_{n,k}^2}} \sqrt{\frac{\widehat{\xi}_{n,k}}{(1+\widehat{\xi}_{n,k})}} . Y_{n,k} \\ &= \sqrt{\frac{\xi_{n,k}^2}{0.667 + \xi_{n,k}^2}} G_{n,k}^{PE} . Y_{n,k} \end{aligned} \tag{2.38}$$

This method will be called parametric spectral power estimation (PARA) in the following sections and chapters. $\xi_{n,k}$ is calculated by using Eq. (2.14).

However, to reduce spectral distortion that results form this subtraction rule the authors [20] proposed the inclusion of a spectral floor

$$|\bar{X}_{n,k}| = \begin{cases} |\widehat{X}_{n,k}|, & \text{if } |\widehat{X}_{n,k}| > \mu|Y_{n,k}| \\ f(\mu, |Y_{n,k}|), & \text{otherwise} \end{cases} \tag{2.39}$$

where $f(\mu, |Y_{n,k}|) = 0.5(\mu|Y_{n,k}| + |\bar{X}_{n-1,k}|)$ and now $|\bar{X}_{n,k}|$ is the estimate for discrete Cosine transformed spectral magnitude of clean speech. In the above subtraction rule, the value of μ to be $0.05 \sim 0.2$.

In many works [15], [20], [30] including all described above authors estimate the *a priori* SNR ($\xi_{n,k}$) using a constant value for the smoothing parameter α Eq. (2.14). The value is usually chosen in the range 0.96 to 0.995.

2.2.2 Spectral amplitude estimation based on Wiener filtering

In this method, a multiplicative filter is assumed and its output is the estimation of the clean signal. Let $W_{n,k}$ denotes the filter gain and $Y_{n,k}$ denotes the noisy speech spectral component. Then an estimate of the clean speech spectral component is obtained as

$$\widehat{X}_{n,k} = W_{n,k} Y_{n,k} \tag{2.40}$$

$W_{n,k}$ is derived in the MMSE sense by minimizing the cost function

$$N_{n,k} = E\{(\widehat{X}_{n,k} - X_{n,k})^2\} \tag{2.41}$$

Substituting Eq. (2.40) into Eq. (2.41)

$$N_{n,k} = E\{(W_{n,k}Y_{n,k} - X_{n,k})^2\} \tag{2.42}$$

Substituting Eq. (2.2) into Eq. (2.42)

$$\begin{aligned} N_{n,k} &= E\{[W_{n,k}(X_{n,k} + D_{n,k}) - X_{n,k}]^2\} \\ &= (W_{n,k}^2 - 2W_{n,k} + 1)E\{X_{n,k}^2\} + 2W_{n,k}(W_{n,k} - 1)E\{X_{n,k}D_{n,k}\} \\ &\quad + W_{n,k}^2 E\{D_{n,k}^2\} \end{aligned} \tag{2.43}$$

Using the fact that $X_{n,k}$ and $D_{n,k}$ are zero-mean and uncorrelated real Gaussian random variables (i.e., $E\{X_{n,k}D_{n,k}\} = 0$, $E\{X_{n,k}\} = 0$ and $E\{D_{n,k}\} = 0$), the above cost function takes the form

$$N_{n,k} = (W_{n,k}^2 - 2W_{n,k} + 1)E\{X_{n,k}^2\} + W_{n,k}^2 E\{D_{n,k}^2\} \tag{2.44}$$

Differentiating $N_{n,k}$ with respect to $W_{n,k}$ gives

$$\begin{aligned}\frac{\partial N_{n,k}}{\partial W_{n,k}} &= (2W_{n,k}-2)E\{X_{n,k}^2\}+2W_{n,k}E\{D_{n,k}^2\} \\ &= 2(W_{n,k}-1)E\{X_{n,k}^2\}+2W_{n,k}E\{D_{n,k}^2\} \end{aligned} \tag{2.45}$$

Equating $\partial N_{n,k}/\partial W_{n,k}$ to zero yields

$$2(W_{n,k}-1)E\{X_{n,k}^2\}+2W_{n,k}E\{D_{n,k}^2\}=0 \tag{2.46}$$

Dividing Eq. (2.46) by $2E\{D_{n,k}^2\}$

$$(W_{n,k}-1)\frac{E\{X_{n,k}^2\}}{E\{D_{n,k}^2\}}+W_{n,k}=0 \tag{2.47}$$

Substituting $E\{X_{n,k}^2\}/E\{D_{n,k}^2\}=\xi_{n,k}$ (defined in Eq. (2.13)) into Eq. (2.47)

$$\begin{aligned}(W_{n,k}-1)\xi_{n,k}+W_{n,k} &= W_{n,k}\xi_{n,k}-\xi_{n,k}+W_{n,k} \\ &= 0 \end{aligned} \tag{2.48}$$

Thus an expression for optimum $W_{n,k}$ is obtained as

$$W_{n,k}=\frac{\xi_{n,k}}{\xi_{n,k}+1} \tag{2.49}$$

This is the gain function of the Wiener filter (Eq. (2.49)) whose output is the estimate of clean speech spectral component and input is the noisy signal spectral component. $\xi_{n,k}$ is calculated by using Eq. (2.14).

2.2.3 Dual gain Wiener filter

The dual gain Wiener filter is proposed by Soon and Koh [43]. The derivation of the gain function for the conventional single gain Wiener filter (Eq. (2.49)) is based on the assumption that the clean speech spectral component $X_{n,k}$ and the noise spectral component $D_{n,k}$ are uncorrelated real Gaussian random variables (i.e., $E\{X_{n,k}D_{n,k}\}=0$). The gain function resulted due to this assumption is only attenuative, i.e., less than 1. But noise can increase or decrease a clean speech spectral component amplitude. The component that has been increased by noise should be decreased by the filter gain and the component that has been decreased by noise should be increased by the filter gain. For the first case filter gain should be less than 1 and for the second case the filter gain should be greater than 1. The derivation of the dual gain Wiener filter [43]is given in the following.

If $X_{n,k}D_{n,k} > 0$

$$E\{X_{n,k}D_{n,k}\} = E\{|X_{n,k}||D_{n,k}|\} \tag{2.50}$$

and if $X_{n,k}D_{n,k} < 0$

$$E\{X_{n,k}D_{n,k}\} = -E\{|X_{n,k}||D_{n,k}|\} \tag{2.51}$$

instead of assuming $E\{X_{n,k}D_{n,k}\} = 0$. The value of $E\{|X_{n,k}|\}$ and $E\{|D_{n,k}|\}$ is calculated as follows.

The probability density function of a random variable x which follows Gaussian distribution is

$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left(-(x-\mu)^2/2\sigma_x^2\right), -\infty < x < \infty \tag{2.52}$$

where σ and μ are the mean and standard deviation of x, respectively. The expected value of x with the distribution given above is defined as

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \tag{2.53}$$

As $X_{n,k}$ is also a random variable and zero-mean ($\mu = 0$), the probability density function of $X_{n,k}$

$$f(X) = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-X^2/2\sigma_X^2\right), -\infty < X < \infty \tag{2.54}$$

But the probability density function of $|X_{n,k}|$ is required. A fundamental theorem on probability density function of a random variable y, when $y = g(x)$, is

$$f_y(y) = \frac{f_x(x_1)}{|g'(x_1)|} + \cdots + \frac{f_x(x_n)}{|g'(x_n)|} + \cdots \tag{2.55}$$

where $g'(x)$ is the derivative of $g(x)$, $f_x(x)$ is the distribution of x. x_1 , $x_2 \cdots$ x_n are the real roots of $y = g(x)$.

In this case, $g(X) = |X|$, $y = g(x)$, i.e., $|X| = g(X)$ has two roots $+X$ and $-X$, $X_1 = +X$, $X_2 = -X$. Using Eq. (2.55)

$$f(|X|) = \frac{f_X(X_1)}{|g'(X_1)|} + \frac{f_X(X_2)}{|g'(X_2)|} \tag{2.56}$$

As $g(X_1) = +X$ therefore $g'(X_1) = +1$ and $|g'(X_1)| = 1$. Similarly $g(X_2) = -X$ therefore $g'(X_2) = -1$ and $|g'(X_2)| = 1$. Also

$$\begin{aligned} f_X(X_1) &= \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-(+X)^2/2\sigma_X^2\right) \\ &= \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left(-X^2/2\sigma_X^2\right) \end{aligned} \tag{2.57}$$

$$\begin{aligned} f_X(X_2) &= \frac{1}{\sigma_X\sqrt{2\pi}}\exp\left(-(-X)^2/2\sigma_X^2\right) \\ &= \frac{1}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right) \end{aligned} \tag{2.58}$$

Substituting Eqs. (2.57) and (2.58) into Eq. (2.56)

$$\begin{aligned} f(|X|) &= \frac{\frac{1}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right)}{1} + \frac{\frac{1}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right)}{1} \\ &= \frac{1}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right) + \frac{1}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right) \\ &= \frac{2}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right) \end{aligned} \tag{2.59}$$

The probability density function $f(|X|)$ is obtained as

$$f(|X|) = \frac{2}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right) \tag{2.60}$$

With the distribution function given in Eq. (2.60) and the definition of expected value given in Eq. (2.53), we get

$$\begin{aligned} E(|X|) &= \int_0^\infty X f(|X|)dX \\ &= \int_0^\infty X\frac{2}{\sigma_X\sqrt{2\pi}}\exp\left(-X^2/2\sigma_X^2\right)dX \\ &= \frac{2}{\sigma_X\sqrt{2\pi}}\int_0^\infty X\exp\left(-X^2/2\sigma_X^2\right)dX \end{aligned} \tag{2.61}$$

Using Eq. (2.33), we obtain

$$\int_0^\infty X\exp\left(-X^2/2\sigma_X^2\right)dX = \frac{\Gamma(\frac{1+1}{2})}{2(1/2\sigma_X^2)^{\frac{1+1}{2}}} \tag{2.62}$$

Thus $E\{|X|\}$ is obtained as

$$\begin{aligned} E(|X|) &= \frac{2}{\sigma_X\sqrt{2\pi}}\frac{\Gamma(\frac{1+1}{2})}{2(1/2\sigma_X^2)^{\frac{1+1}{2}}} \\ &= \frac{2}{\sigma_X\sqrt{2\pi}}\frac{\Gamma(1)}{2(1/2\sigma_X^2)^1} \\ &= \frac{2}{\sigma_X\sqrt{2\pi}}\frac{1}{2(1/2\sigma_X^2)} \\ &= \sqrt{\frac{2}{\pi}}\sigma_X \end{aligned} \tag{2.63}$$

i.e.,

$$E\{|X_{n,k}|\} = \sqrt{\frac{2}{\pi}}\sigma_X(k) \tag{2.64}$$

Similarly,

$$E(|D_{n,k}|) = \sqrt{\frac{2}{\pi}}\sigma_D(k) \tag{2.65}$$

Finally, substituting Eqs. (2.64) and (2.65) into Eqs. (2.50) and (2.51), respectively,

$$E\{X_{n,k}D_{n,k}\} = \frac{2}{\pi}\sigma_X(k)\sigma_D(k), \quad \text{if} \quad X_{n,k}D_{n,k} > 0 \tag{2.66}$$

and

$$E\{X_{n,k}D_{n,k}\} = -\frac{2}{\pi}\sigma_X(k)\sigma_D(k), \quad \text{if} \quad X_{n,k}D_{n,k} < 0 \tag{2.67}$$

where $\sigma_X(k)$ and $\sigma_D(k)$ are the standard deviations of the clean speech spectral $X_{n,k}$ and the noise spectral component $D_{n,k}$, respectively (i.e., $\sigma_X(k)/\sigma_D(k) = \sqrt{\xi_{n,k}}$). Substituting Eq. (2.66) into Eq. (2.43), we obtain

$$\begin{aligned} N_{n,k} &= (W_{n,k}^2 - 2W_{n,k} + 1)E\{X_{n,k}^2\} + 2W_{n,k}(W_{n,k} - 1)E\{X_{n,k}D_{n,k}\} \\ &\quad + W_{n,k}^2 E\{D_{n,k}^2\} \\ &= (W_{n,k}^2 - 2W_{n,k} + 1)E\{X_{n,k}^2\} + 2W_{n,k}(W_{n,k} - 1)\frac{2}{\pi}\sigma_X(k)\sigma_D(k) \\ &\quad + W_{n,k}^2 E\{D_{n,k}^2\} \end{aligned} \tag{2.68}$$

Differentiating $N_{n,k}$ with respect to $W_{n,k}$ gives

$$\begin{aligned} \frac{\partial N_{n,k}}{\partial W_{n,k}} = \quad & (2W_{n,k} - 2)E\{X_{n,k}^2\} + 2(2W_{n,k} - 1)\frac{2}{\pi}\sigma_X(k)\sigma_D(k) \\ & + 2W_{n,k}E\{D_{n,k}^2\} \end{aligned} \tag{2.69}$$

Equating $\partial N_{n,k}/\partial W_{n,k}$ to zero yields

$$(2W_{n,k} - 2)E\{X_{n,k}^2\} + 2(2W_{n,k} - 1)\frac{2}{\pi}\sigma_X(k)\sigma_D(k) + 2W_{n,k}E\{D_{n,k}^2\} = 0 \tag{2.70}$$

Dividing Eq. (2.70) by $2E\{D_{n,k}^2\}$ and substituting $E\{X_{n,k}^2\}/E\{D_{n,k}^2\} = \xi_{n,k}$ (defined in Eq. (2.13))

$$(W_{n,k} - 1)\xi_{n,k} + (2W_{n,k} - 1)\frac{\frac{2}{\pi}\sigma_X(k)\sigma_D(k)}{E\{D_{n,k}^2\}} + W_{n,k} = 0 \tag{2.71}$$

Substituting $E\{D_{n,k}^2\} = \sigma_D^2(k)$ as in Eqs. (2.12) and (2.13) into Eq. (2.71), we obtain

$$(W_{n,k} - 1)\xi_{n,k} + (2W_{n,k} - 1)\frac{2}{\pi}\frac{\sigma_X(k)}{\sigma_D(k)} + W_{n,k} = (W_{n,k} - 1)\dot{\xi}_{n,k}$$

$$
\begin{aligned}
&+ \quad (2W_{n,k}-1)\frac{2}{\pi}\sqrt{\xi_{n,k}} + W_{n,k} \\
&= \quad W_{n,k}(\xi_{n,k}+1+\frac{4}{\pi}\sqrt{\xi_{n,k}}) \\
&\quad -\xi_{n,k} - \frac{2}{\pi}\sqrt{\xi_{n,k}} \qquad (2.72)
\end{aligned}
$$

Rearranging Eq. (2.72), the optimum filter gain

$$
\begin{aligned}
G_1 &= W_{n,k} \\
&= \frac{\xi_{n,k} + \frac{2}{\pi}\sqrt{\xi_{n,k}}}{\xi_{n,k}+1+\frac{4}{\pi}\sqrt{\xi_{n,k}}} \qquad (2.73)
\end{aligned}
$$

G_1 is always less than 1 and, the authors [43] have proposed to use this gain for the spectral component whose magnitude has been increased by noise, i.e., for the condition $X_{n,k}D_{n,k} > 0$.

Similarly, substituting Eq. (2.67) into Eq. (2.43) and equating $\partial N_{n,k}/\partial W_{n,k}$ to zero gives the optimum filter gain

$$
\begin{aligned}
G_2 &= W_{n,k} \\
&= \frac{\xi_{n,k} - \frac{2}{\pi}\sqrt{\xi_{n,k}}}{\xi_{n,k}+1-\frac{4}{\pi}\sqrt{\xi_{n,k}}} \qquad (2.74)
\end{aligned}
$$

The authors [43] have proposed to use G_2 for the spectral component whose magnitude has been reduced by noise, i.e., for the condition $X_{n,k}D_{n,k} < 0$.

2.3 New Constraint for Dual Gain Wiener Filter

The dual gain Wiener filter described in the last section requires a new constraint, otherwise G_2 defined in Eq. (2.74) cannot be always guaranteed to be > 1. The value of the G_2 will be > 1 if

$$\xi_{n,k} - \frac{2}{\pi}\sqrt{\xi_{n,k}} > \xi_{n,k} + 1 - \frac{4}{\pi}\sqrt{\xi_{n,k}} \qquad (2.75)$$

i.e.,

$$\xi_{n,k} - \frac{2}{\pi}\sqrt{\xi_{n,k}} - \xi_{n,k} - 1 + \frac{4}{\pi}\sqrt{\xi_{n,k}} = \frac{2}{\pi}\sqrt{\xi_{n,k}} - 1 > 0 \qquad (2.76)$$

which yields

$$\xi_{n,k} > \frac{\pi^2}{4} \qquad (2.77)$$

If $\xi_{n,k} > \frac{\pi^2}{4}$ for a noisy speech spectral component then we propose to use the gain function G_2, otherwise no denoising operation is performed, i.e., $\widehat{X}_{n,k} = Y_{n,k}$.

2.4 Conclusion

In this Chapter, the basic speech enhancement algorithms have been discussed elaborately in the DCT domain. In particular, the methods those haveb een covered are spectral power subtraction (PE), modified spectral power subtraction (MPE), parametric spectral power subtraction (PARA), Wiener filtering and dual gain Wiener filtering. It has been observed that all described methods require the knowledge of the *a priori* SNR. Estimation of the *a priori* SNR requires an averaging parameter according to the "decision-directed" approach. In the next Chapter, an optimal smoothing parameter to estimate the *a priori* SNR in the DCT domain is proposed.

Chapter 3

Enhancement in the DCT Domain using Optimal Estimate of the *a priori* SNR

3.1 Introduction

Various speech-processing systems have found their way in our everyday life through their vivid use in voice communication, speech and speaker recognition, aid for hearing impaired and numerous other applications [2]. However, in many practical situations they are confronted with high ambient noise levels and their performance degrades drastically. Thus, their is a strong need to improve the performance of these systems in noisy conditions by developing speech enhancement algorithms that are able to work at very low SNRs. For single channel speech enhancement, the spectral subtraction based algorithms are commonly used. However, the spectral subtraction algorithms have a serious drawback in that the enhanced speech is accompanied by unpleasant musical noise artifact, which is characterized by tones with random frequencies [1]. Apart from being extremely annoying to the listeners, the musical noise also hampers the performance of the speech-coding algorithms to a great extent [42]. It has been shown in [17]that the key point behind the reduction of *musical noise* by the minimum-mean-squared-error (MMSE) estimator [15] is the use of *a priori* SNR. Several methods such as spectral subtraction based algorithms, Wiener filtering require the knowledge of the *a priori* SNR and the estimation of *a priori* SNR using the "decision-directed" approach requires an averaging parameter [17]. A low value of the averaging parameter is suitable for rapidly changing speech regions, while

a high value is suitable for near stationary speech frames [43]. Conventionally, a constant value is used for the averaging parameter [30].

In this chapter of this research work, we propose and derive an optimal averaging parameter to estimate the a *priori* SNR using the "decision-directed" approach in the DCT domain.

3.2 Problem Formulation

Noise always corrupts speech and is unavoidable in real life. Usually, the noise $d(t)$ is modelled as an additive Gaussian process with zero-mean and variance σ_d^2. The noisy speech signal $y(t)$ is then given by

$$y(t) = x(t) + d(t) \tag{3.1}$$

where $x(t)$ is the clean speech signal. The forward DCT of the noisy signal $\{y(t), 0 \leq t \leq N-1\}$ is given by [30]

$$Y_k = \alpha_k \sum_{t=0}^{N-1} y(t) \cos\left[\frac{\pi(2t+1)k}{2N}\right], 0 \leq k \leq N-1 \tag{3.2}$$

where

$$\alpha_k = \begin{cases} \sqrt{\frac{1}{N}}, k = 0 \\ \sqrt{\frac{2}{N}}, 1 \leq k \leq N-1 \end{cases} \tag{3.3}$$

The objective of this research is to denoise the speech signal with enhanced SNR and better subjective performance by modifying the noisy DCT coefficients Y_k using spectral subtraction techniques incorporating our proposed parameter. The reconstructed signal, $\widehat{x}(t)$, can be obtained using the following inverse Cosine transformation (IDCT) [30]

$$\widehat{x}(t) = \sum_{k=0}^{N-1} \alpha_k \widehat{X}_k \cos\left[\frac{\pi(2t+1)k}{2N}\right], 0 \leq t \leq N-1 \tag{3.4}$$

where $\widehat{X}_k$ denotes the denoised DCT coefficients.

3.3 Adaptive Averaging Parameter for *a priori* SNR Estimation

It was previously mentioned in Chapter 2 that in many works [15], [20], [30] to estimate the *a priori* SNR, α was set to a constant value (e.g., 0.96 to 0.995).

But using a constant α has certain drawbacks. Consider an example as a test case where $\alpha = 0.98$ and the SNR_{post} shows a pulse like behavior, i.e., for $n < n_1$ and $n > n_2$, it is very low as compared to its values in the interval $n_1 \sim n_2$, where n_1 and n_2 denote, respectively, the frames of rising and falling edges. At n_1, a signal component suddenly goes high such that $\gamma_{n_1,k} >> \gamma_{n_1-1,k}$. Since $\widehat{\xi}_{n,k}$ contains 98% of the previous frame estimated SNR, it will fail to respond to this change. Rather $\widehat{\xi}_{n,k}$ will rise slowly and ultimately begin to follow SNR_{post} in this high SNR region ($n_1 \sim n_2$) with some delay. Similarly at n_2, $\widehat{\xi}_{n,k}$ fails to respond to the abrupt downfall of SNR_{post} and only after a certain delay converges to the low SNR level. Therefore, it will be logical to use a much smaller value of α in these transitional areas. This suggests us to use a time-frequency varying α.

The proposed modification in the estimation of *a priori* SNR is given by

$$\widehat{\xi}_{n,k}^{p} = \alpha_{n,k}\tilde{\xi}_{n-1,k} + (1-\alpha_{n,k})P[\gamma_{n,k}-1] \tag{3.5}$$

where $\tilde{\xi}_{n-1,k} = \widehat{X}_{n-1,k}^2/\widehat{\sigma}_d^2(n-1,k)$ and $\alpha_{n,k}$ denotes the time-frequency varying averaging parameter. Now the PE gain function (Eq. (2.15)) can be written as

$$[G_{n,k}^{PE})]^p = \sqrt{\frac{\widehat{\xi}_{n,k}^{p}}{1+\widehat{\xi}_{n.k}^{p}}} \tag{3.6}$$

The MPE gain function (Eq. (2.23)) is given by

$$[G_{n,k}^{MPE}]^p = \frac{1-q_{n,k}}{1-q_{n,k}+q_{n,k}\sqrt{(1+\widehat{\xi}_{n,k}^{p})}\exp(-\nu_{n,k})}[G_{n,k}^{PE}]^p \tag{3.7}$$

where $\nu_{n,k} = \gamma_{n,k}\widehat{\xi}_{n,k}^{p}/\left(2(1+\widehat{\xi}_{n,k}^{p})\right)$. The PARA gain function (Eq. (2.38)) is given by

$$[G_{n,k}^{PARA}]^p = \sqrt{\frac{(\widehat{\xi}_{n,k}^{p})^2}{0.667+(\widehat{\xi}_{n,k}^{p})^2}}[G_{n,k}^{PE}]^p \tag{3.8}$$

The Wiener gain function (Eq. (2.49)) is given by

$$W_{n,k}^{p} = \frac{\widehat{\xi}_{n,k}^{p}}{\widehat{\xi}_{n,k}^{p}+1} \tag{3.9}$$

The gain functions of dual gain Wiener filter are given by

$$W_{n,k}^{p} = \begin{cases} \dfrac{\widehat{\xi}_{n,k}^{p}+\frac{2}{\pi}\sqrt{\widehat{\xi}_{n,k}^{p}}}{\widehat{\xi}_{n,k}^{p}+1+\frac{4}{\pi}\sqrt{\widehat{\xi}_{n,k}^{p}}}, & \text{if} \quad X_{n,k}D_{n,k} > 0 \\ \dfrac{\widehat{\xi}_{n,k}^{p}-\frac{2}{\pi}\sqrt{\widehat{\xi}_{n,k}^{p}}}{\widehat{\xi}_{n,k}^{p}+1-\frac{4}{\pi}\sqrt{\widehat{\xi}_{n,k}^{p}}}, & \text{if} \quad X_{n,k}D_{n,k} < 0 \end{cases} \tag{3.10}$$

3.3.1 Estimation of adaptive averaging parameter

It is desired that the estimate $\widehat{\xi}^{p}_{n,k}$ given by Eq. (3.5) should actually be as close as possible to *a priori* SNR $\xi_{n,k}$, we propose an MMSE estimator for $\alpha_{n,k}$ which minimizes the error

$$\begin{aligned} J_{\alpha} &= E\left\{(\widehat{\xi}^{p}_{n,k} - \xi_{n,k})^2 \mid \tilde{\xi}_{n-1,k}\right\} \\ &= E\left\{(\widehat{\xi}^{p}_{n,k})^2 - 2\widehat{\xi}^{p}_{n,k}\xi_{n,k} + \xi^2_{n,k}) \mid \tilde{\xi}_{n-1,k}\right\} \end{aligned} \tag{3.11}$$

given $\tilde{\xi}_{n-1,k}$. The operator $P[\cdot]$ in Eq. (3.5) is used to ensure the positiveness of the decision-directed estimator in case $\gamma_{n,k} - 1$ goes negative. As stated in [15] it is also possible to apply $P[\cdot]$ on the right side of Eq. (3.5) rather than using only on $\gamma_{n,k} - 1$. In both the case results are very similar [14]. Thus as the objective of using $P[\cdot]$ is to ensure only positiveness of $\widehat{\xi}_{n,k}$, not because as a part of basic mathematical derivation, we omit temporarily the operator $P[\cdot]$ in the definition of our cost function to facilitate taking the expectation operator. Substituting Eq. (3.5) into Eq. (3.11), we obtain

$$\begin{aligned} J_{\alpha} &= E\Big\{\big(\alpha^2_{n,k}\tilde{\xi}^2_{n-1,k} + (1-\alpha_{n,k})^2(\gamma_{n,k}-1)^2 \\ &\quad +2\alpha_{n,k}(1-\alpha_{n,k})\tilde{\xi}_{n-1,k}(\gamma_{n,k}-1) \\ &\quad -2\xi_{n,k}[\alpha_{n,k}\tilde{\xi}_{n-1,k} + (1-\alpha_{n,k})(\gamma_{n,k}-1)] + \xi^2_{n,k}\big) \mid \tilde{\xi}_{n-1,k}\Big\} \\ &= \alpha^2_{n,k}\tilde{\xi}^2_{n-1,k} + (1-\alpha_{n,k})^2 E\left\{(\gamma_{n,k}-1)^2\right\} \\ &\quad +2\alpha_{n,k}(1-\alpha_{n,k})\tilde{\xi}_{n-1,k}E\left\{(\gamma_{n,k}-1)\right\} \\ &\quad -2\alpha_{n,k}\xi_{n,k}\tilde{\xi}_{n-1,k} - 2(1-\alpha_{n,k})\xi_{n,k}E\left\{(\gamma_{n,k}-1)\right\} + \xi^2_{n,k} \end{aligned} \tag{3.12}$$

Using Eq. (2.12) we can write

$$\begin{aligned} E\{(\gamma_{n,k}-1)^2\} &= E\{\gamma^2_{n,k} - 2\gamma_{n,k} + 1\} \\ &= \frac{E\{Y^4_{n,k}\}}{E\{D^2_{n,k}\}^2} - 2E\{\gamma_{n,k}\} + 1 \end{aligned} \tag{3.13}$$

To find the value of $E\{(\gamma_{n,k}-1)^2\}$ we need to evaluate $E\{Y^4_{n,k}\}$. For notational simplicity we drop subscript (n,k) as in Chapter 2. Using Eq. (2.2), an expression for Y^2 can be obtained as

$$Y^2 = (X^2 + 2XD + D^2) \tag{3.14}$$

Thus $E\{Y^4\}$ is obtained as

$$\begin{aligned} E\{Y^4\} &= E\{X^4 + 4X^2D^2 + D^4 + 2X^3D \\ &\quad +2XD^3 + 2X^2D^2\} \\ &= E\{X^4\} + 4E\{X^2\}E\{D^2\} + E\{D^4\} \\ &\quad +2E\{X^3\}E\{D\} + 2E\{X\}E\{D^3\} \\ &\quad +2E\{X^2\}E\{D^2\} \end{aligned} \tag{3.15}$$

Now again using the assumption for $X_{n,k}$ and $D_{n,k}$ (i.e., zero-mean and uncorrelated real Gaussian random variables), the simplified form of $E\{Y^4\}$ can be obtained as

$$\begin{aligned} E\{Y^4\} &= E\{X^4\} + 4E\{X^2\}E\{D^2\} + E\{D^4\} + 2E\{X^2\}E\{D^2\} \\ &= E\{X^4\} + E\{D^4\} + 6E\{X^2\}E\{D^2\} \end{aligned} \tag{3.16}$$

Again introducing notational subscript (n,k)

$$\begin{aligned} E\{Y_{n,k}^4\} &= E\{X_{n,k}^4\} + E\{D_{n,k}^4\} \\ &\quad +6E\{X_{n,k}^2\}E\{D_{n,k}^2\} \end{aligned} \tag{3.17}$$

Using Eqs. (2.31)-(2.36)

$$E\{X_{n,k}^4\} = 3E\{X_{n,k}^2\}^2 \tag{3.18}$$

Similarly,

$$E\{D_{n,k}^4\} = 3E\{D_{n,k}^2\}^2 \tag{3.19}$$

Substituting Eqs. (3.18) and (3.19) into Eq. (3.17)

$$\begin{aligned} E\{Y_{n,k}^4\} &= 3E\{X_{n,k}^2\}^2 + 3E\{D_{n,k}^2\}^2 \\ &\quad +6E\{X_{n,k}^2\}E\{D_{n,k}^2\} \end{aligned} \tag{3.20}$$

Using Eqs. (3.13) and (3.20), we obtain

$$\begin{aligned} E\{(\gamma_{n,k}-1)^2\} &= \frac{E\{Y_{n,k}^4\}}{E\{D_{n,k}^2\}^2} - 2E\{\gamma_{n,k}\} + 1 \\ &= \frac{3E\{X_{n,k}^2\} + 3E\{D_{n,k}^2\} + 6E\{X_{n,k}^2\}E\{D_{n,k}^2\}}{E\{D_{n,k}^2\}^2} \\ &\quad -2E\{\gamma_{n,k}\} + 1 \\ &= 3\frac{E\{X_{n,k}^2\}^2}{E\{D_{n,k}^2\}^2} + 3\frac{E\{D_{n,k}^2\}^2}{E\{D_{n,k}^2\}^2} + 6\frac{E\{X_{n,k}^2\}E\{D_{n,k}^2\}}{E\{D_{n,k}^2\}^2} \\ &\quad -2E\{\gamma_{n,k}\} + 1 \\ &= 3\xi_{n,k}^2 + 3 + 6\xi_{n,k} - 2E\{\gamma_{n,k}\} + 1 \end{aligned} \tag{3.21}$$

As $E\{(\gamma_{n,k}-1)\} = \xi_{n,k}$, it follows that $E\{\gamma_{n,k}\} = 1+\xi_{n,k}$ [[15], Eq. (49)].So Eq. (3.21) becomes

$$\begin{aligned} E\{(\gamma_{n,k}-1)^2\} &= 3\xi_{n,k}^2 + 6\xi_{n,k} + 3 - 2(1+\xi_{n,k}) + 1 \\ &= 3\xi_{n,k}^2 + 4\xi_{n,k} + 2 \end{aligned} \tag{3.22}$$

Substituting $E\{(\gamma_{n,k}-1)\} = \xi_{n,k}$ and $E\{(\gamma_{n,k}-1)^2\} = 3\xi_{n,k}^2 + 4\xi_{n,k} + 2$ into Eq. (3.12), we obtain

$$\begin{aligned} J_\alpha &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + (1-\alpha_{n,k})^2(3\xi_{n,k}^2 + 4\xi_{n,k} + 2) + 2\alpha_{n,k}(1-\alpha_{n,k})\tilde{\xi}_{n-1,k}\xi_{n,k} \\ &\quad -2\alpha_{n,k}\xi_{n,k}\tilde{\xi}_{n-1,k} - 2(1-\alpha_{n,k})\xi_{n,k}\xi_{n,k} + \xi_{n,k}^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + 3\xi_{n,k}^2(1-\alpha_{n,k})^2 + 4\xi_{n,k}(1-\alpha_{n,k})^2 + 2(1-\alpha_{n,k})^2 \\ &\quad +2\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}\xi_{n,k}\tilde{\xi}_{n-1,k} \\ &\quad -2(1-\alpha_{n,k})\xi_{n,k}\xi_{n,k} + \xi_{n,k}^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + 3\xi_{n,k}^2(1-\alpha_{n,k})^2 + 4\xi_{n,k}(1-\alpha_{n,k})^2 + 2(1-\alpha_{n,k})^2 \\ &\quad +2\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} - 2(1-\alpha_{n,k})\xi_{n,k}^2 \\ &\quad +\xi_{n,k}^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \xi_{n,k}^2\left[3(1-\alpha_{n,k})^2 - 2(1-\alpha_{n,k}) + 1\right] \\ &\quad +\xi_{n,k}\left[4(1-\alpha_{n,k})^2 + 2\alpha_{n,k}\tilde{\xi}_{n-1,k} - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k} - 2\alpha_{n,k}\tilde{\xi}_{n-1,k}\right] \\ &\quad +2(1-\alpha_{n,k})^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \xi_{n,k}^2(3 - 6\alpha_{n,k} + 3\alpha_{n,k}^2 - 2 + 2\alpha_{n,k} + 1) + \xi_{n,k}\left[4(1-\alpha_{n,k})^2\right. \\ &\quad \left.-2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\right] + 2(1-\alpha_{n,k})^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \xi_{n,k}^2(2 - 4\alpha_{n,k} + 3\alpha_{n,k}^2) + \xi_{n,k}\left[4(1-\alpha_{n,k})^2 - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\right] \\ &\quad +2(1-\alpha_{n,k})^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \alpha_{n,k}^2\xi_{n,k}^2 + \xi_{n,k}^2(2 - 4\alpha_{n,k} + 2\alpha_{n,k}^2) - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} \\ &\quad +4\xi_{n,k}(1-\alpha_{n,k})^2 + 2(1-\alpha_{n,k})^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \alpha_{n,k}^2\xi_{n,k}^2 + 2\xi_{n,k}^2(1-\alpha_{n,k})^2 - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} \\ &\quad +4\xi_{n,k}(1-\alpha_{n,k})^2 + 2(1-\alpha_{n,k})^2 \\ &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} + \alpha_{n,k}^2\xi_{n,k}^2 + 2\xi_{n,k}^2(1-\alpha_{n,k})^2 \\ &\quad +4\xi_{n,k}(1-\alpha_{n,k})^2 + 2(1-\alpha_{n,k})^2 \\ &= \alpha_{n,k}^2(\tilde{\xi}_{n-1,k}^2 - 2\tilde{\xi}_{n-1,k}\xi_{n,k} + \xi_{n,k}^2) + 2(1-\alpha_{n,k})^2(\xi_{n,k}^2 + 2\xi_{n,k} + 1) \\ &= \alpha_{n,k}^2(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 2(1-\alpha_{n,k})^2(\xi_{n,k}+1)^2 \end{aligned} \tag{3.23}$$

Differentiating J_α with respect to $\alpha_{n,k}$ gives

$$\begin{aligned}\frac{\partial J_\alpha}{\partial \alpha_{n,k}} &= 2\alpha_{n,k}(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 4(1-\alpha_{n,k})(-1)(\xi_{n,k}+1)^2 \\ &= 2\alpha_{n,k}(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 4\alpha_{n,k}(\xi_{n,k}+1)^2 - 4(\xi_{n,k}+1)^2 \\ &= \alpha_{n,k}\left\{2(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 4(\xi_{n,k}+1)^2\right\} - 4(\xi_{n,k}+1)^2 \qquad (3.24)\end{aligned}$$

Now equating $\partial J_\alpha / \partial \alpha_{n,k}$ to zero yields

$$\alpha_{n,k}\left\{2(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 4(\xi_{n,k}+1)^2) - 4(\xi_{n,k}+1\right\}^2 = 0 \qquad (3.25)$$

Finally, the optimum expression for $\alpha_{n,k}$ is obtained as

$$\alpha_{n,k}^{opt} = \frac{1}{1 + 0.5\left(\frac{\xi_{n,k} - \tilde{\xi}_{n-1,k}}{\xi_{n,k}+1}\right)^2} \qquad (3.26)$$

3.3.2 Implementation of $\alpha_{n,k}^{opt}$

As $\xi_{n,k}$ is unknown, Eq. (3.26) cannot be used directly. Nevertheless, an approximate value of $\alpha_{n,k}$ can be obtained substituting $\bar{\xi}_{n,k} = E\{\gamma_{n,k} - 1\}$ for $\xi_{n,k}$ in Eq. (3.26). This is a reasonable substitution as $E\{\bar{\xi}_{n,k}\} \cong \xi_{n,k}$. If SNR_{post} over a region shows uniform variation, $\alpha_{n,k}$ will attain a value close to 1. For any abrupt change $\alpha_{n,k}$ attains a lower value enabling $\hat{\xi}_{n,k}^{p}$ to respond to that change more suitably.

3.4 Results

3.4.1 Data used

For evaluating the impact of the proposed time-frequency varying smoothing parameter on the traditional spectral subtraction methods i.e., PE, MPE, PARA, Wiener filter and dual gain Wiener filter, simulations were performed over a data set consisting of 20 different speech utterances from the TIMIT and other sources. Half of the sentences are spoken by female speakers while the remaining sentences are by male speakers. The speech signals are sampled at 8 KHz and quantized to 16 bits. Noise types in our experiments were also from NOISEX database and they were white Gaussian, Babble, Aircockpit, Helicockpit and Highway. The results are shown for one female speech utterance- "Pretty soon a woman

came with along with a folded umbrella as a walking stick" (S1) and male speech utterance- "She had your dark suit in greasy wash water all year" (S2) with noise cases- white Gaussian, Babble, Highway and Aircockpit in following section and chapter.

3.4.2 Estimation of noise level

Noise is estimated from noise signal itself. Noise is estimated according to [41]

$$[\widehat{\sigma}_d(n,k)]^{\beta_N} = \lambda_D[\widehat{\sigma}_d(n-1,k)]^{\beta_N} + (1-\lambda_D)|N_{n,k}|^{\beta_N} \tag{3.27}$$

where $0.5 \leq \lambda_D \leq 0.9$ and $|N_{n,k}|$ is noise spectral's magnitude. In this thesis work, we have used $\lambda_D = 0.9$ and $\beta_N = 2$ for all cases.

3.4.3 Performance test

The frame-basis analysis is performed in all cases. The frame is of 32 ms. That is each time, we have taken 256 samples of noisy speech. Input frames are taken as 75 percent overlapped. To reconstruct signal, weighted overlap-add method is used. To quantify the performance of the proposed smoothing parameter on the conventional spectral subtraction rules and the Wiener filter, IS (Itakura-Saito) Distortion, average segmental SNR (AvgSegSNR) and overall output SNR (Output SNR) are measured. Both the IS measures and AvgSegSNRs show high correlation with informal listening tests. The performance evaluation techniques that we have used are given below briefly.

3.4.3.1 IS distortion measure

For an original clean frame of speech with linear prediction (LP) coefficient vector, $\vec{a}_\phi$, and processed speech coefficient vector, $\vec{a}_d$, the Itakura-Saito distortion measure defined by [46] is given by,

$$d_{IS}(\vec{a}_d\vec{a}_\phi) = \left[\frac{\sigma_\phi^2}{\sigma_d^2}\right]\left[\frac{\vec{a}_d R\phi\vec{a}_d^T}{\vec{a}_\phi R\phi\vec{a}_\phi^T}\right] + \log\left(\frac{\sigma_d^2}{\sigma_\phi^2}\right) - 1 \tag{3.28}$$

where σ_d^2 and σ_ϕ^2 represent the all-pole gains for the processed and clean speech frame respectively. The lower the IS measure for an enhanced speech, the better is its perceived quality.

3.4.3.2 AvgSegSNR measure

The frame-based segmental SNR is formed by averaging frame level SNR estimates and is defined by [46]

$$\text{AvgSegSNR} = \frac{10}{M}\sum_{m=0}^{M-1}\log_{10}\frac{\sum_{t=Nm}^{Nm+N-1}x^2(t)}{\sum_{t=Nm}^{Nm+N-1}(x(t)-\widehat{x}(t))^2} \quad \text{dB} \tag{3.29}$$

where M denotes the number of frames, and $\widehat{x}(t)$ may by the reconstructed signal or the noisy signal. The lower and upper thresholds are selected to be -10 dB and 35 dB, respectively. The higher the AvgSegSNR measure for an enhanced speech, the better is its perceived quality.

3.4.3.3 Overall SNR measure

The overall SNR of the noisy signal, i.e., the input SNR is defined as

$$\text{SNR} = 10\log_{10}\frac{\sum_{t=0}^{N-1}x^2(t)}{\sum_{t=0}^{N-1}d^2(t)} \quad \text{dB} \tag{3.30}$$

The output SNR is measured by the same equation with the exception that the noise is now calculated as the difference between the original and enhanced signal, i.e., substituting $d(t)$ by $\widehat{d}(t) = x(t) - \widehat{x}(t)$, where $\widehat{x}(t)$ denotes the estimated speech. The higher the output SNR measure for an enhanced speech, the better is its perceived quality.

3.4.4 Performance evaluation

Various comparative results are shown for the PE, MPE, PARA, Wiener filtering and dual gain Wiener filtering methods using our proposed smoothing parameter $\alpha_{n,k}$ along with using $\alpha = 0.98$ in Figs (3.1)-(3.14).

Figs. (3.1)-(3.4) show the variation of IS measures, AvgSegSNRs and overall SNRs, respectively, at different noise levels and at different noise types for the PE and the MPE methods. Below 5 dB SNRs, the impact of the proposed $\alpha_{n,k}$ is not significant in IS measure, AvgSegSNR and overall SNR. But for SNR levels above 5 dB, improvements in all of the objective measures are noticeable. The PE method using $\alpha = 0.98$ fails to improve the IS measure for input SNR of 15 dB and above for babble noise, and overall SNR for input SNR of 23 dB and above. The MPE method using $\alpha = 0.98$ fails to improve the IS measure for input SNR of 7 dB and above, and overall SNR for input SNR of 15 dB and above for white noise. On the contrary, using $\alpha_{n,k}$ IS measures has improved significantly for the PE method, and AvgSegSNRs and overall SNRs have improved significantly for the MPE method for SNR $\geq$ 15 dB. It is to be noted that for the MPE algorithm $q_{n,k} = 0.2$ is assumed (Eq. (3.7)). This is the value empirically used by Ephraim and Malah [15] and previously by McAullay and Malpass [18]. However, it has been shown in [44] that for power subtraction algorithms using such a constant value as the speech presence probability gives poor results. To overcome this, a time-frequency varying $q_{n,k}$ was developed. In this work, we have used a constant $q_{n,k}$ to emphasize the importance of the self-adaptive $\alpha_{n,k}$.

Figs. (3.5)-(3.8) show the variation of IS measures, AvgSegSNRs and overall SNRs, respectively, at different noise levels and at different noise types for the PARA method. For comparison, IS measures, AvgSegSNRs and overall SNRs of the PE and MPE methods are also presented in these figures. At low SNRs, using the proposed $\alpha_{n,k}$ has more impact on the improvement in IS measure than on the improvement in AvgSegSNR. But at high SNRs, improvements in all of the objective measures are noticeable. Also note that the PARA method using $\alpha = 0.98$ fails to improvethe IS measure for input SNR of 11 dB and above. But the use of the self-adaptive $\alpha_{n,k}$ proposed in this work has ensured significant quality improvement. It has been observed that for a given utterance the performance of the conventional PARA method is highly dependent on the

choice of α. Another point to be noted with this method is that if no 'floor' (Eq. (2.39)) is used, the improvement in SNR is reasonably good but with the cost paid in quality. This is the reason why the 'floor' was introduced even at the cost of overall SNR. It is evident from Figs. (3.5)-(3.8) that the self-adaptive $\alpha_{n,k}$ has further improved the performance of the algorithm in quality terms. Therefore, there is a greater flexibility in changing the 'floor' parameters, and accordingly higher quality at a comparatively higher SNR can be ensured. It can also be concluded that the IS measure is most significant for the PARA method than the MPE and the PE methods; and AvgSegSNR and overall SNR are most significant for the MPE method than the PE and the PARA methods using $\alpha_{n,k}$ for all types of noises.

Tables 3.1-3.3 show comparative IS measures, AvgSegSNRs and overall SNRs for the Wiener filter at different noise levels for S1 corrupted by Gaussian white noise. As can be seen, the use of proposed $\alpha_{n,k}$ improves the quality in terms of IS, AvgSegSNRs and overall SNRS of the enhanced speech significantly. The effectiveness of $\alpha_{n,k}$ on the Wiener filter is also shown in Figs. (3.9)-(3.12) for highway and aircockpit noises. For comparison, SNRs of the MPE method and IS measures of the PARA method are also presented in these figures. The Wiener filter using $\alpha = 0.98$ fails to improve the IS measure for the whole range of input SNR while it fails to improve AvgSegSNR and overall SNR for input SNR of 15 dB and above. The proposed $\alpha_{n,k}$ has more impact on the improvement in AvgSegSNR and overall SNR than on the improvement in IS measure, particularly at SNRs 10 dB and above. Below 10 dB SNRs, proposed $\alpha_{n,k}$ has more impact on IS measure. Better quality of enhanced speech in terms of AvgSegSNR and overall SNR is obtained by the Wiener filter than that of the MPE method.

In Figs. (3.13)-(3.14), the effect of $\alpha_{n,k}$ on the dual gain Wiener filter is shown. The same impact as in the Wiener filter is observed on the dual gain Wiener filter. The theoretical limits of IS measures, AvgSegSNRs and overall SNRs for the dual gain Wiener filter are shown in Figs. (3.15)-(3.16) and in Table 3.4. In this case, clean signal is used for decision making, i.e., to separate which noisy speech spectral component require gain G_1 and which noisy speech spectral component require gain G_2 (Eq. (2.73) and (2.74)). Theoretical limits of all objective measures are well ahead for the constraint dual gain Wiener filter

indicating a potential scope for further research.

Speech enhancement results in the time and frequency domain are also shown in Fig. 3.17. Fig. 3.17 (a) shows the white noise degraded speech $y(t)$ at SNR = 10 dB for the female utterance S1 ("Pretty soon a woman came along with a folded umbrella as a walking stick"), and the enhanced speech resulting from the MPE method using $\alpha = 0.98$ and the proposed $\alpha_{n,k}$. Fig. 3.17 (b) shows the corresponding spectrograms. It is apparent that the speech energy in the MPE method using proposed $\alpha_{n,k}$ is maintained, unlike the conventional MPE which reduces the speech energy.

Table 3.1: Results on IS improvement for the speech utterance "Pretty soon a woman came along with a folded umbrella as a walking stick", corrupted by additive white noise at different SNRs

SNR	IS		
dB	Degraded	Wiener using $\alpha = 0.98$	Wiener using $\alpha_{n,k}$
−5	4.4329	30.7794	3.1740
0	3.7006	54.4804	3.1953
5	2.9595	63.6960	2.9944
10	2.2565	60.9298	2.3838
15	1.6316	53.4421	1.9824
20	1.0988	45.3573	1.7358
30	0.4169	9.7157	0.9950

Table 3.2: Results on AvgSegSNR improvement for the speech utterance "Pretty soon a woman came along with a folded umbrella as a walking stick", corrupted by additive white noise at different SNRs

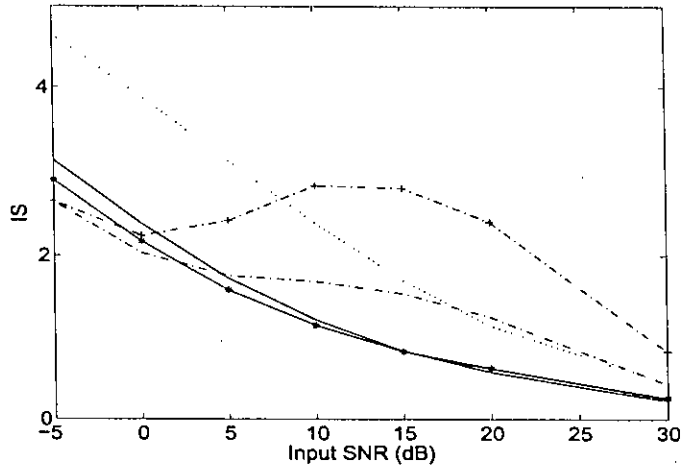
SNR	AvgSegSNR		
dB	Degraded	Wiener using $\alpha = 0.98$	Wiener using $\alpha_{n,k}$
−5	-5.5230	-0.1055	-0.6835
0	-3.1505	1.7793	1.8087
5	-0.2367	3.7671	4.4540
10	3.1788	5.9274	7.4043
15	7.1140	8.3957	10.6785
20	11.4041	11.1947	14.1348
30	20.5644	18.0996	21.7406

Table 3.3: Results on overall SNR improvement for the speech utterance "Pretty soon a woman came along with a folded umbrella as a walking stick", corrupted by additive white noise at different SNRs

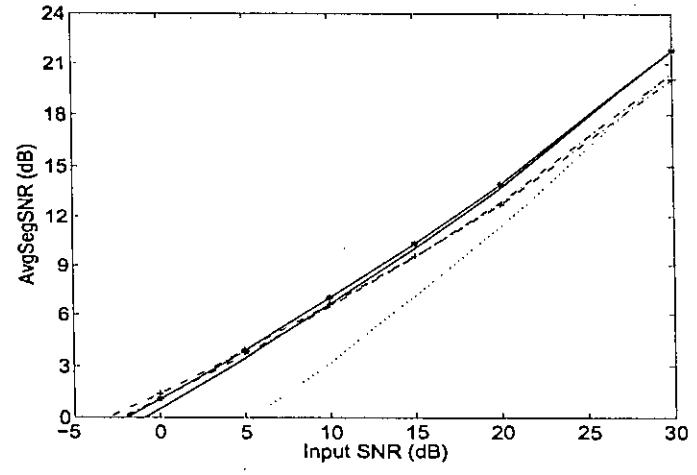
SNR	overall SNR		
dB	Degraded	Wiener using $\alpha = 0.98$	Wiener using $\alpha_{n,k}$
−5	-2.9793	3.6153	5.1304
0	0.9694	5.6284	8.2346
5	5.3421	8.0384	11.3117
10	10.0585	10.8128	14.7315
15	15.0333	13.9526	18.4320
20	20.0190	17.4296	22.2472
30	30.0065	25.5371	30.6844

Table 3.4: Results of theoretical limit on overall SNR improvement for the speech utterance "Pretty soon a woman came along with a folded umbrella as a walking stick", corrupted by additive white noise at different SNRs

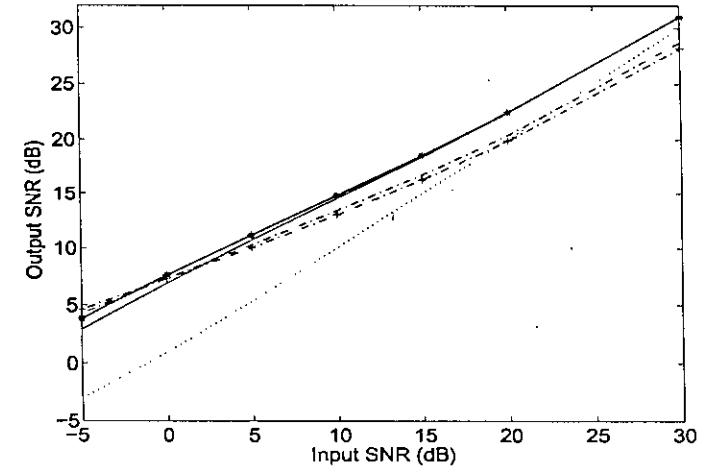
SNR	overall SNR		
dB	Wiener	dual gain Wiener	constraint dual gain
−5	5.1304	4.6550	5.8623
0	8.2346	8.0150	9.8219
5	11.3117	11.5971	13.7349
10	14.7315	15.4374	17.7062
15	18.4320	19.2670	21.8034
20	22.2472	23.3977	26.0932
30	30.6844	32.8261	35.4797



(a)

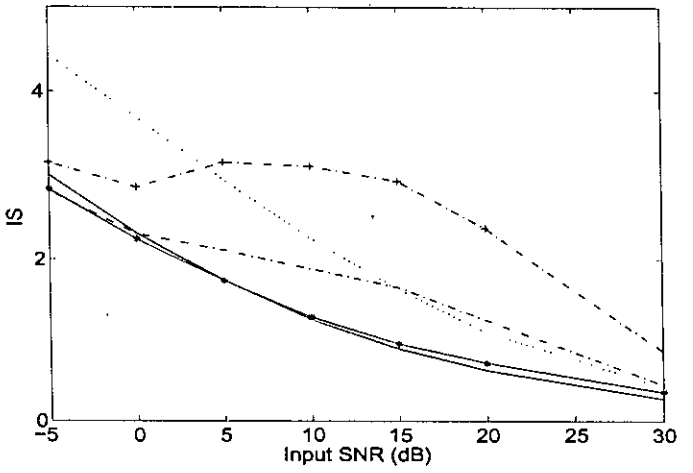


(b)

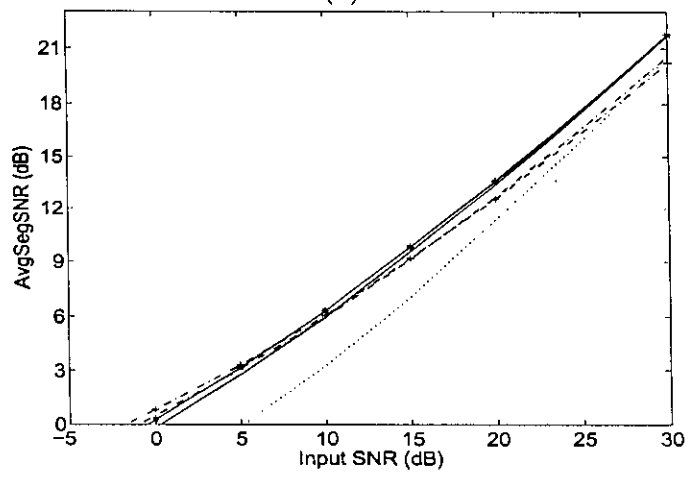


(c)

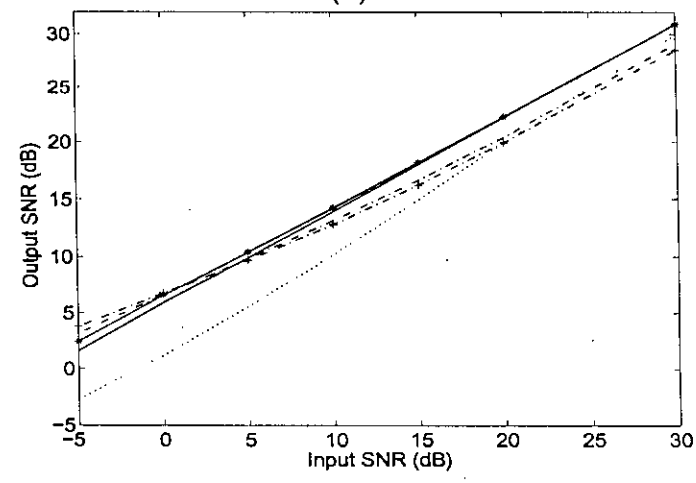
Fig. 3.1: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE and MPE where (...) for degraded, (−.) for PE using $\alpha = 0.98$, (-) for PE using $\alpha_{n,k}$, (+−.) for MPE using $\alpha = 0.98$ and (∗−) for MPE using $\alpha_{n,k}$ (S1, white noise).



(a)



(b)



(c)

Fig. 3.2: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE and MPE where (...) for degraded, $(-.)$ for PE using $\alpha = 0.98$, (-) for PE using $\alpha_{n,k}$, $(+-.)$ for MPE using $\alpha = 0.98$ and $(*-)$ for MPE using $\alpha_{n,k}$ (S1, babble noise).

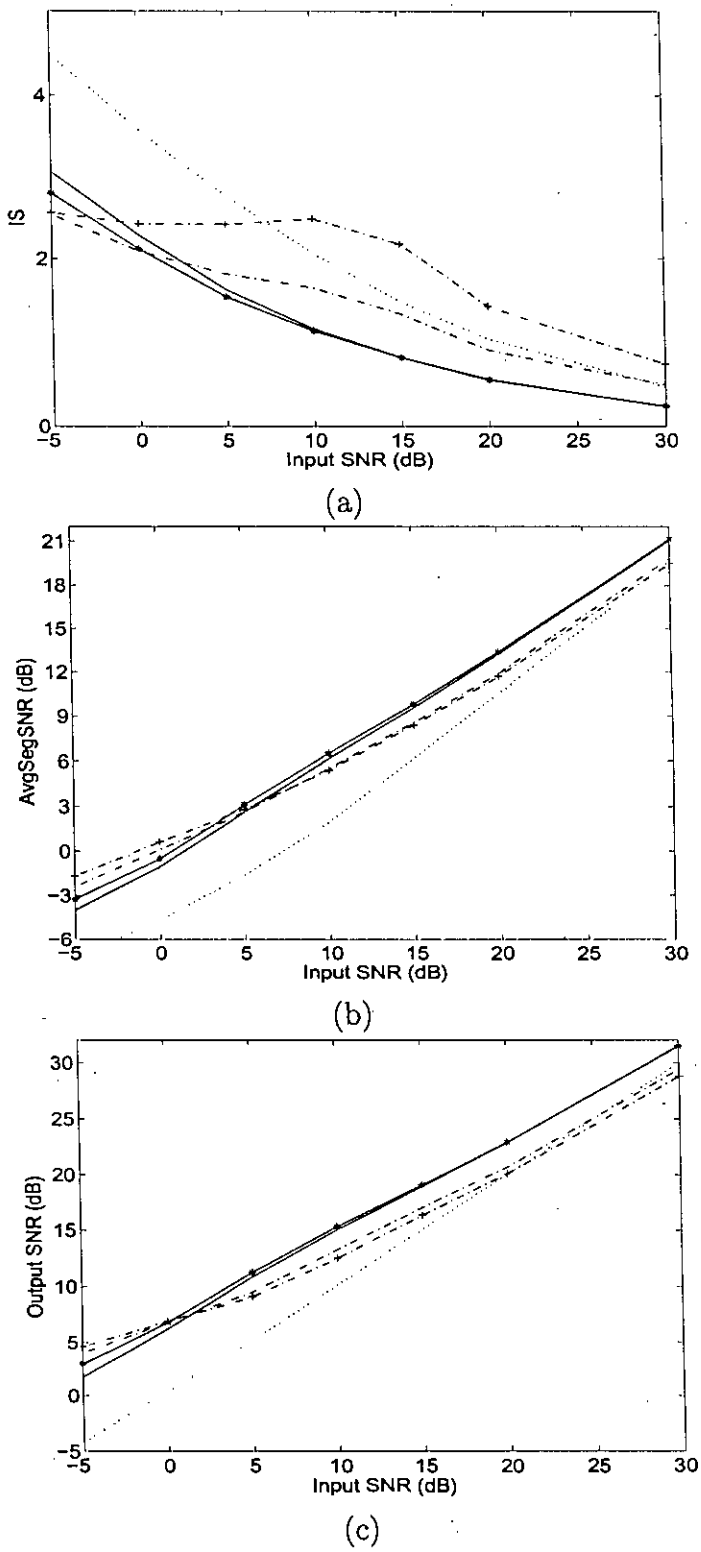


Fig. 3.3: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE and MPE where (...) for degraded, $(-.)$ for PE using $\alpha = 0.98$, (-) for PE using $\alpha_{n,k}$, $(+-.)$ for MPE using $\alpha = 0.98$ and $(*-)$ for MPE using $\alpha_{n,k}$ (S2, highway noise).

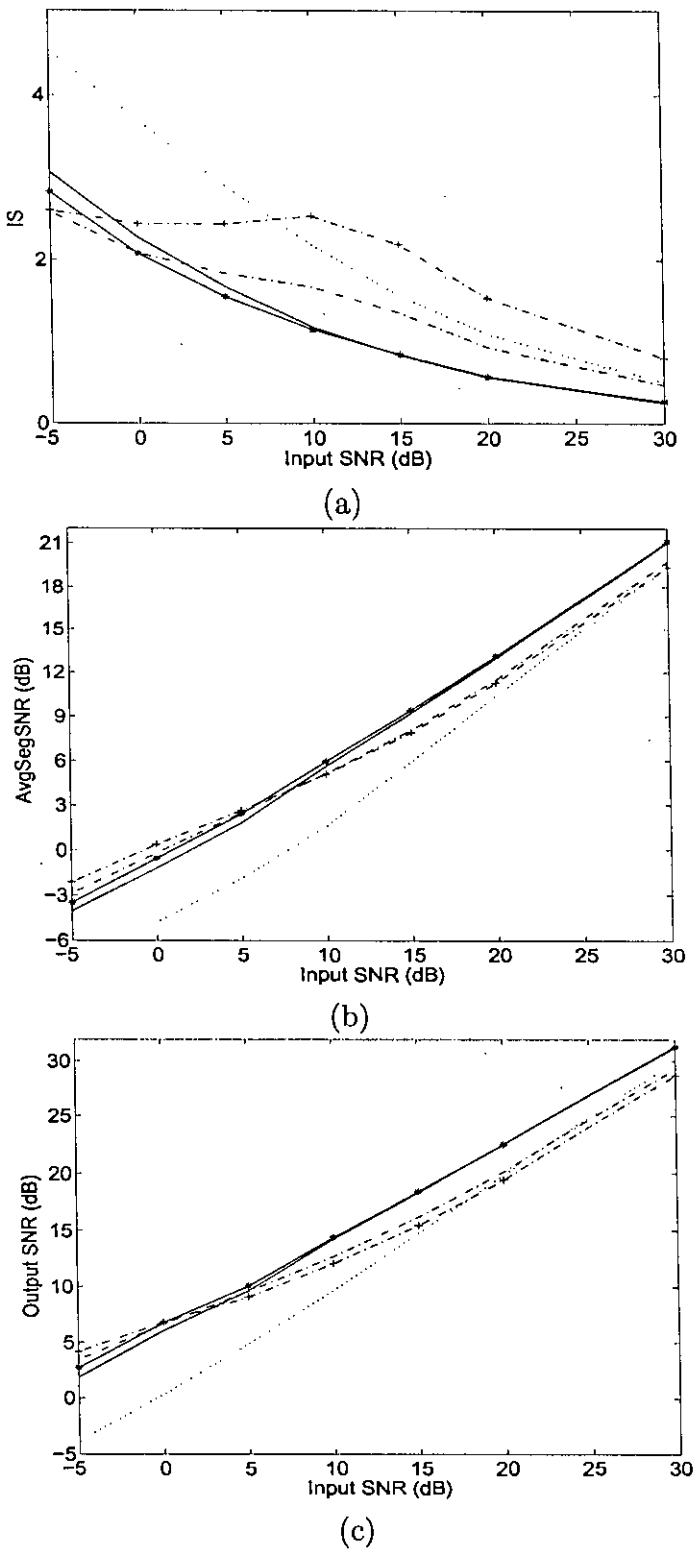
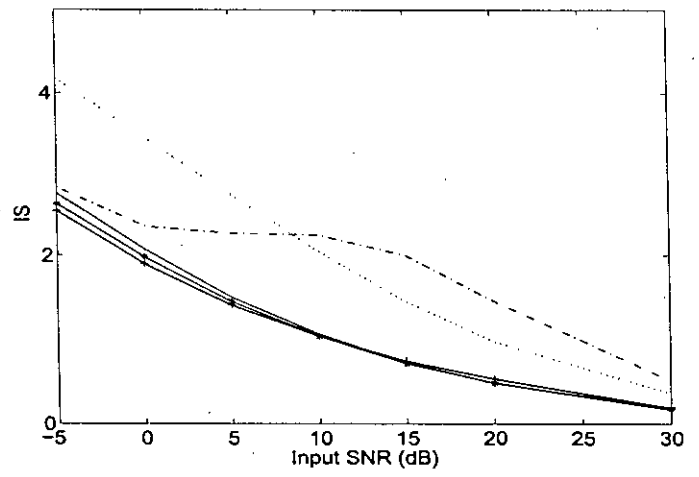
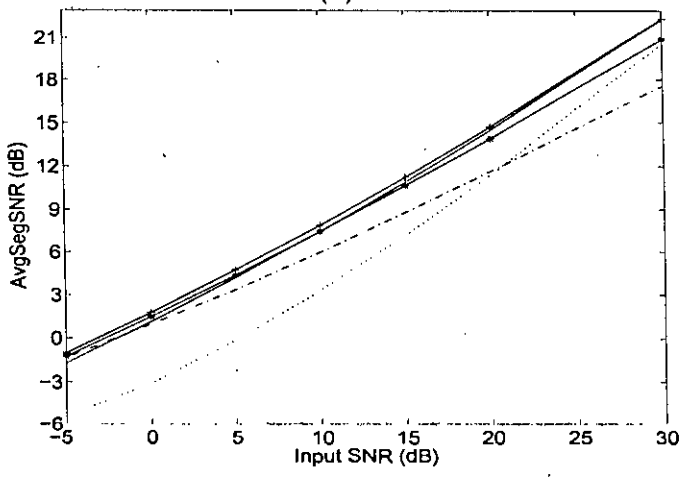


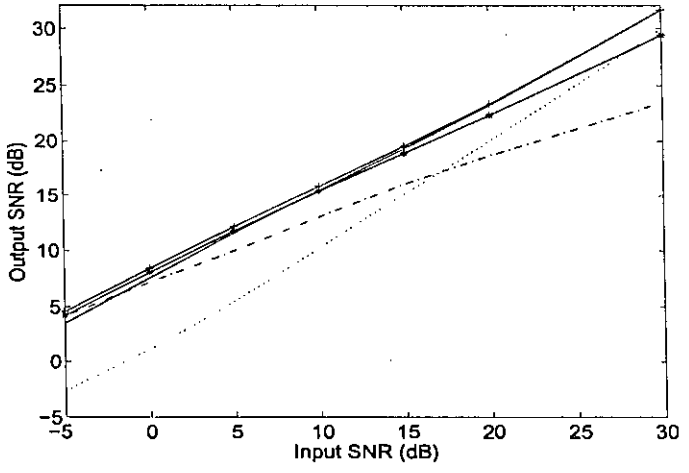
Fig. 3.4: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE and MPE where (...) for degraded, $(-.)$ for PE using $\alpha = 0.98$, (-) for PE using $\alpha_{n,k}$, $(+-.)$ for MPE using $\alpha = 0.98$ and $(*-)$ for MPE using $\alpha_{n,k}$ (S2, aircockpit noise).



(a)

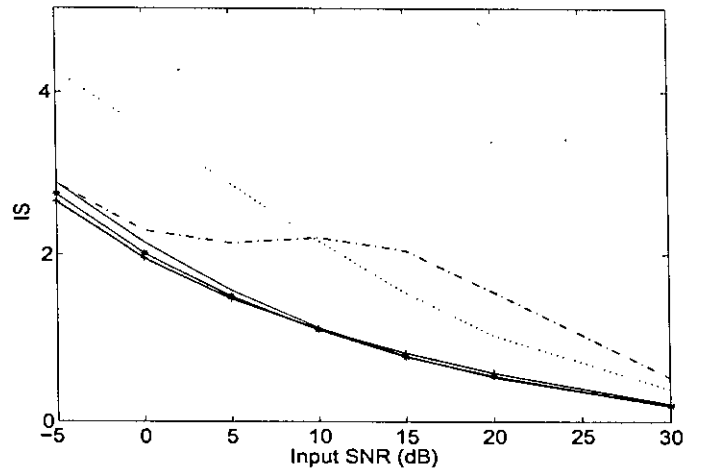


(b)

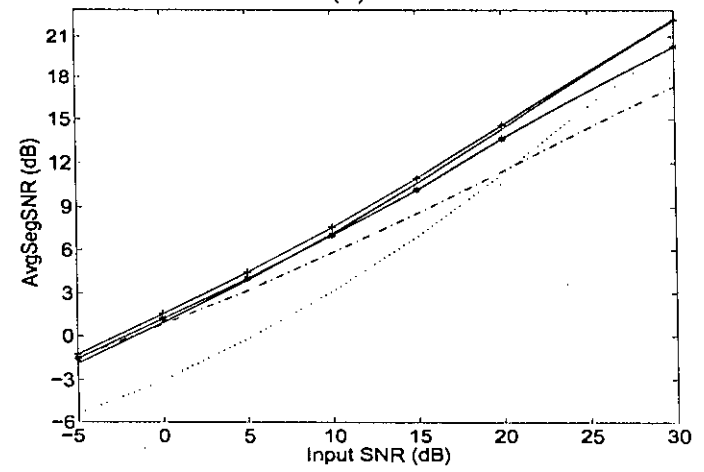


(c)

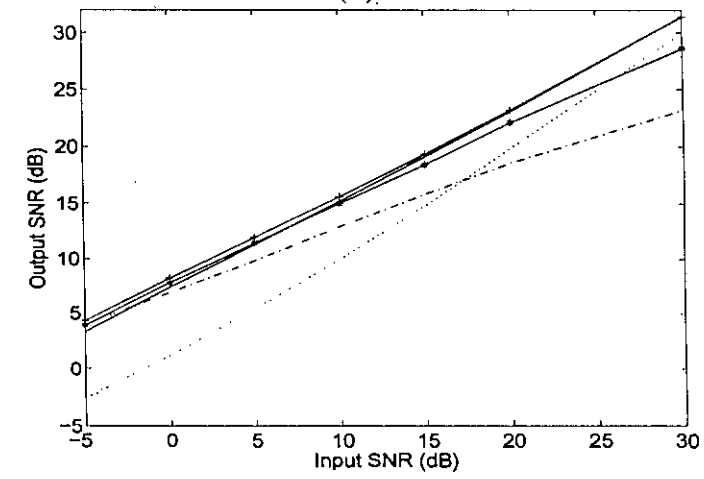
Fig. 3.5: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE, MPE and PARA where (...) for degraded, (-) for PE using $\alpha_{n,k}$, (+−) for MPE using $\alpha_{n,k}$, (−.) for PARA using $\alpha = 0.98$ and (∗−) for PARA using $\alpha_{n,k}$ (S1, highway noise).



(a)



(b)



(c)

Fig. 3.6: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE, MPE and PARA where (...) for degraded, (-) for PE using $\alpha_{n,k}$, (+−) for MPE using $\alpha_{n,k}$, (−.) for PARA using $\alpha = 0.98$ and (∗−) for PARA using $\alpha_{n,k}$ (S1, aircockpit noise).

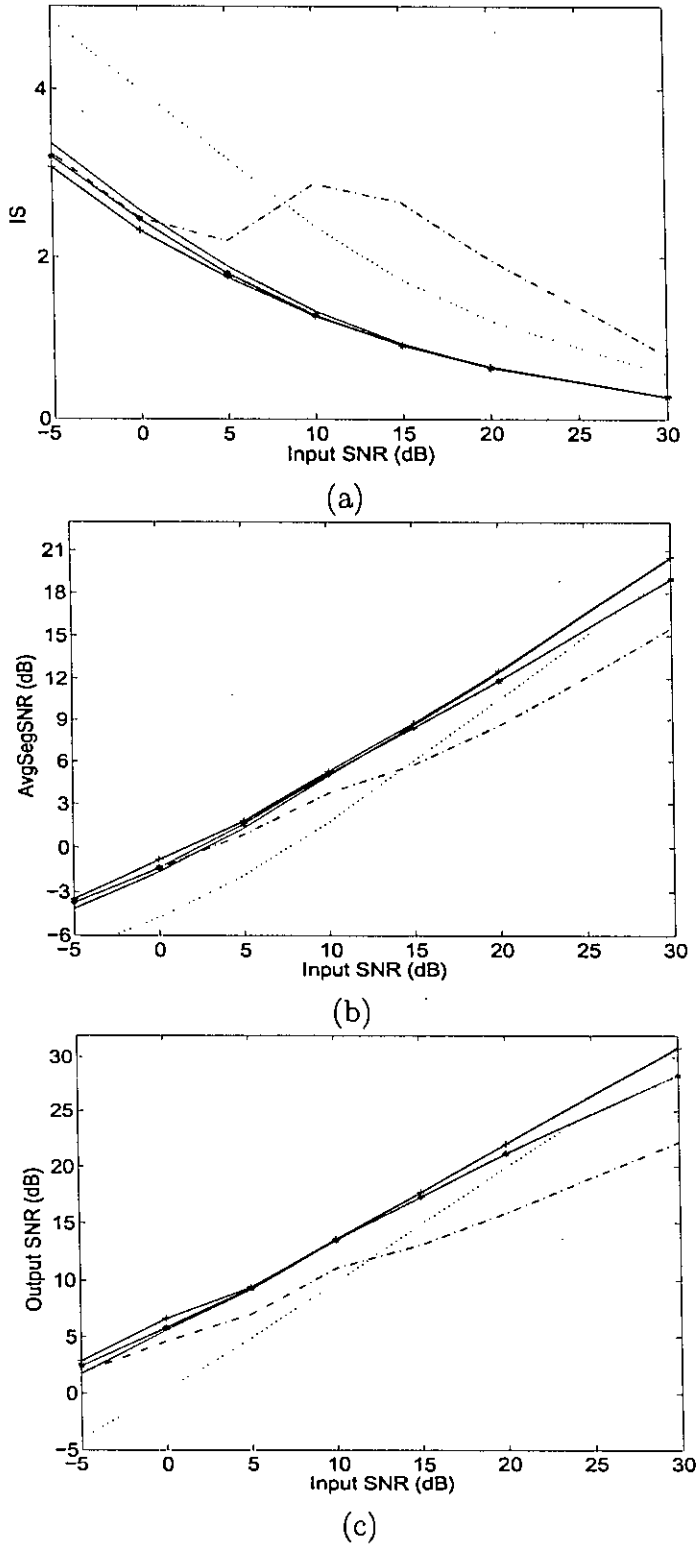


Fig. 3.7: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE, MPE and PARA where (...) for degraded, (-) for PE using $\alpha_{n,k}$, (+−) for MPE using $\alpha_{n,k}$, (−.) for PARA using $\alpha = 0.98$ and (∗−) for PARA using $\alpha_{n,k}$ (S2, white noise).

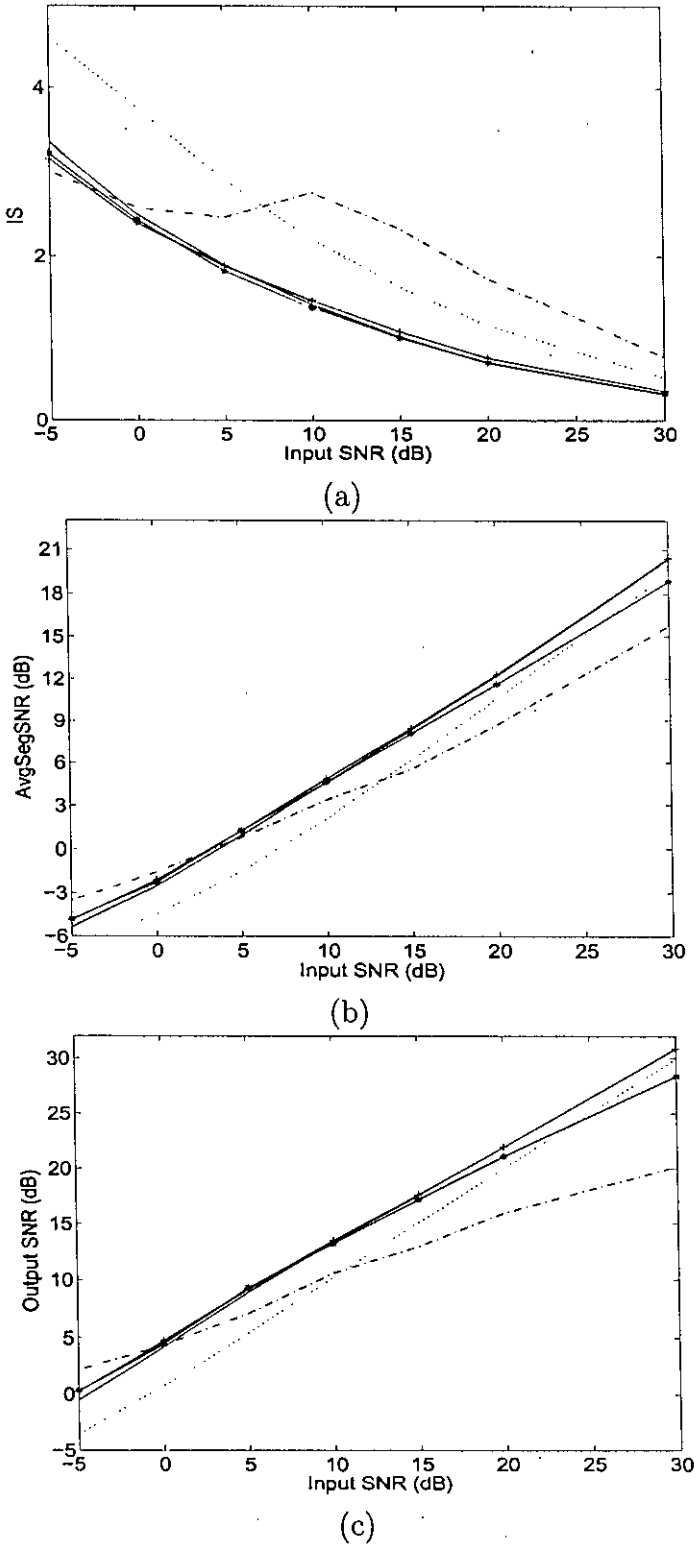


Fig. 3.8: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PE, MPE and PARA where (...) for degraded, (-) for PE using $\alpha_{n,k}$, (+−) for MPE using $\alpha_{n,k}$, (−.) for PARA using $\alpha = 0.98$ and (∗−) for PARA using $\alpha_{n,k}$ (S2, babble noise).

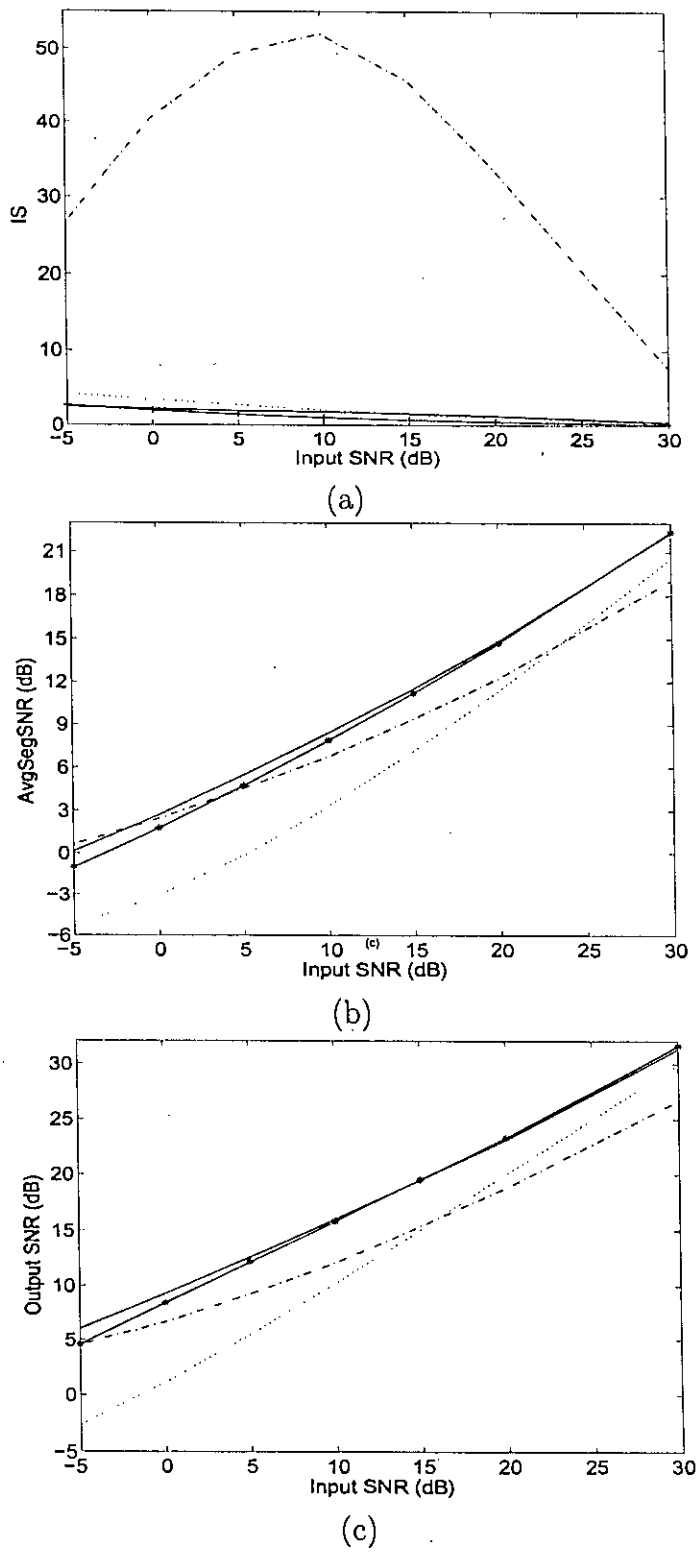
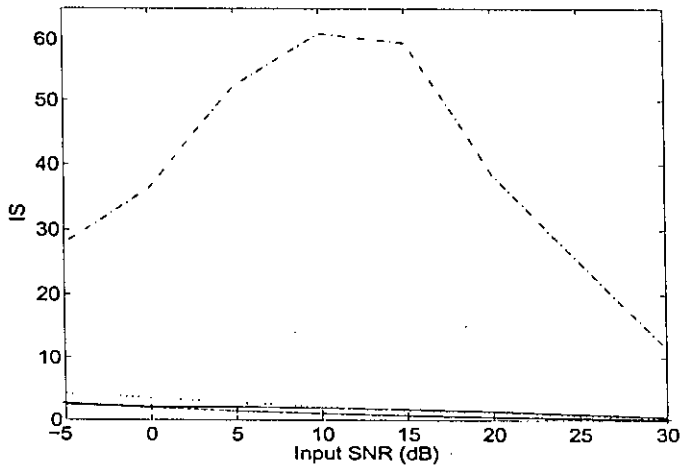
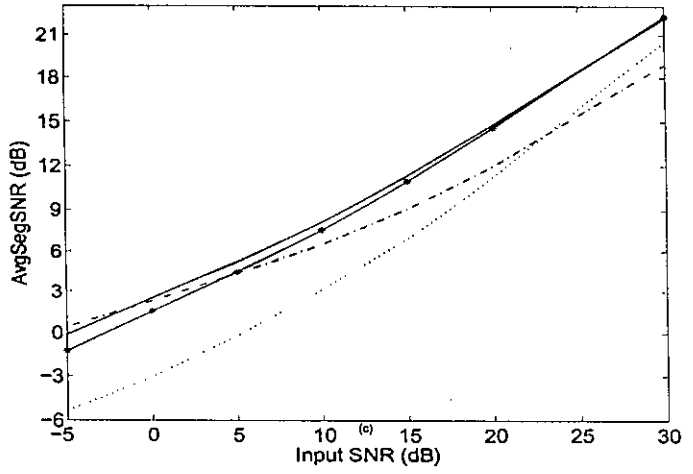


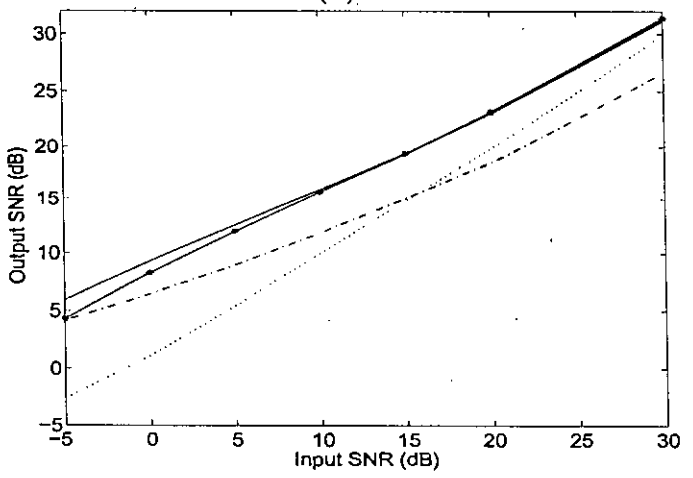
Fig. 3.9: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA and Wiener (...) for degraded, (+−) for PARA using $\alpha_{n,k}$, (∗−) for MPE using $\alpha_{n,k}$, (−.) for Wiener filter using $\alpha = 0.98$ and (-) for Wiener filter using $\alpha_{n,k}$ (S1, highway noise).



(a)

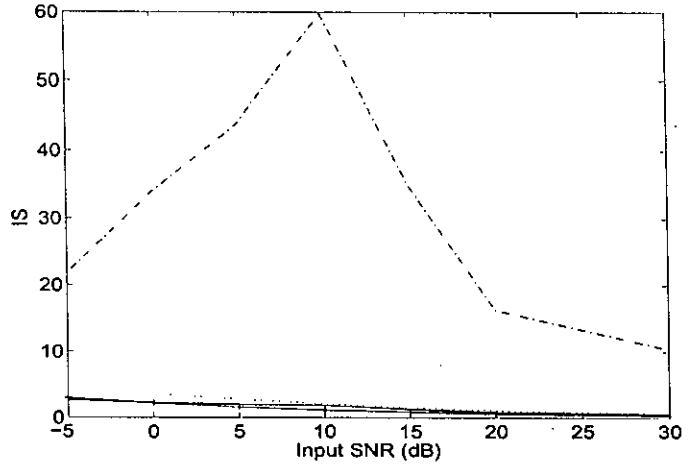


(b)

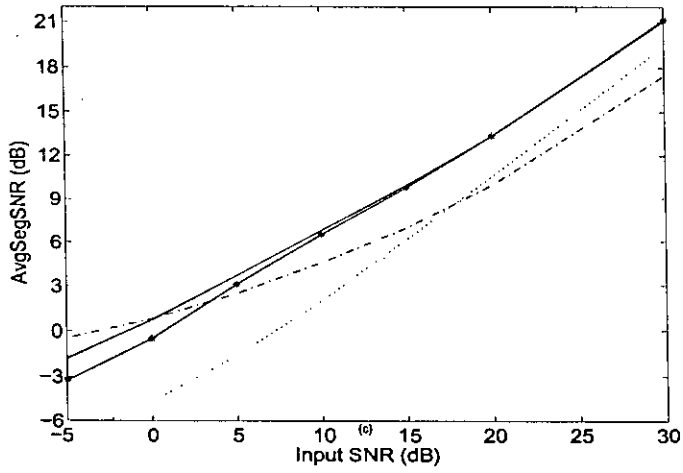


(c)

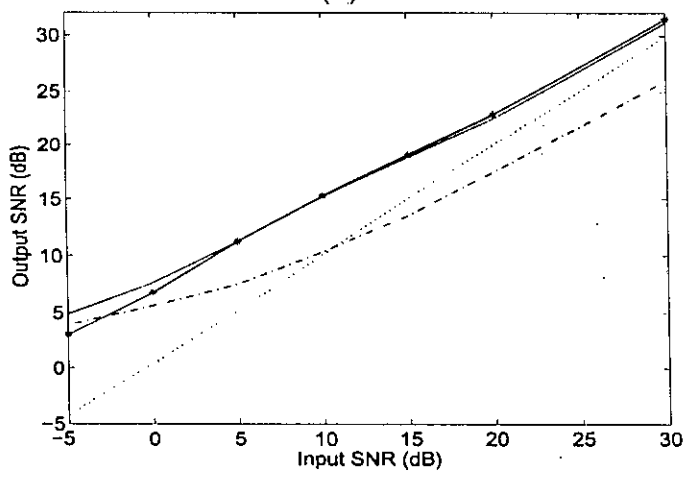
Fig. 3.10: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA and Wiener (...) for degraded, (+−) for PARA using $\alpha_{n,k}$, (∗−) for MPE using $\alpha_{n,k}$, (−.) for Wiener filter using $\alpha = 0.98$ and (-) for Wiener filter using $\alpha_{n,k}$ (S1, aircockpit noise).



(a)

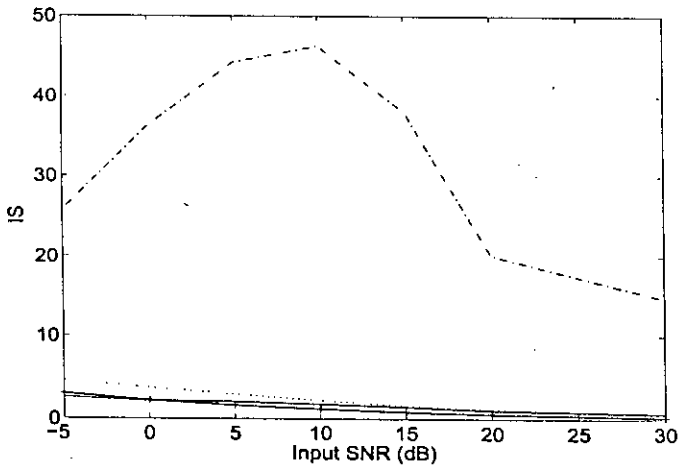


(b)

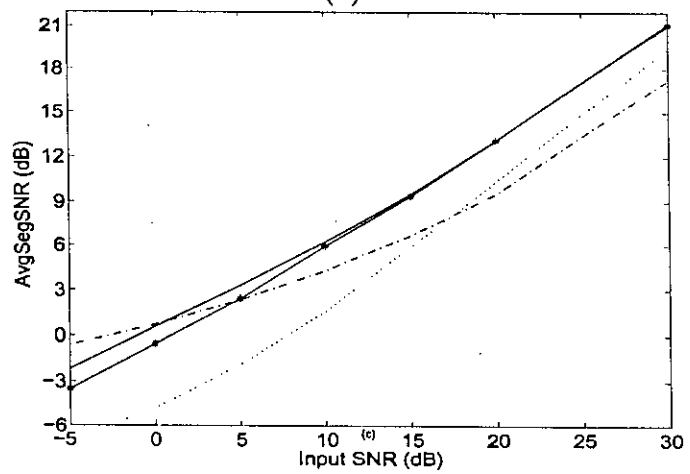


(c)

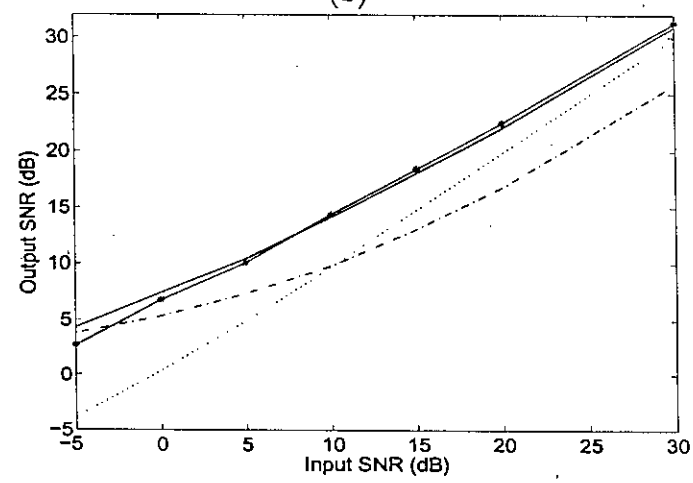
Fig. 3.11: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA and Wiener (...) for degraded, (+−) for PARA using $\alpha_{n,k}$, (∗−) for MPE using $\alpha_{n,k}$, (−.) for Wiener filter using $\alpha = 0.98$ and (-) for Wiener filter using $\alpha_{n,k}$ (S2, highway noise).



(a)

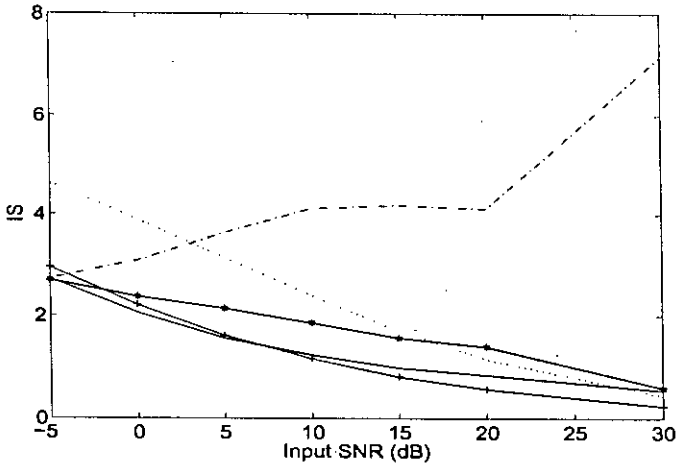


(b)

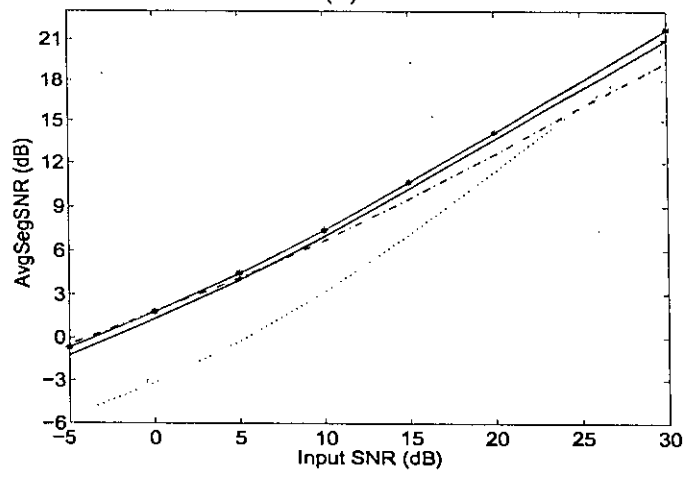


(c)

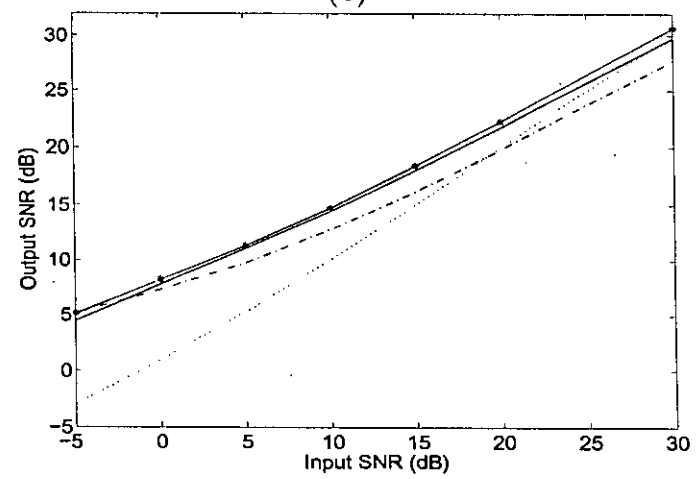
Fig. 3.12: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA and Wiener (...) for degraded, (+−) for PARA using $\alpha_{n,k}$, (∗−) for MPE using $\alpha_{n,k}$, (−.) for Wiener filter using $\alpha = 0.98$ and (-) for Wiener filter using $\alpha_{n,k}$ (S2, aircockpit noise).



(a)

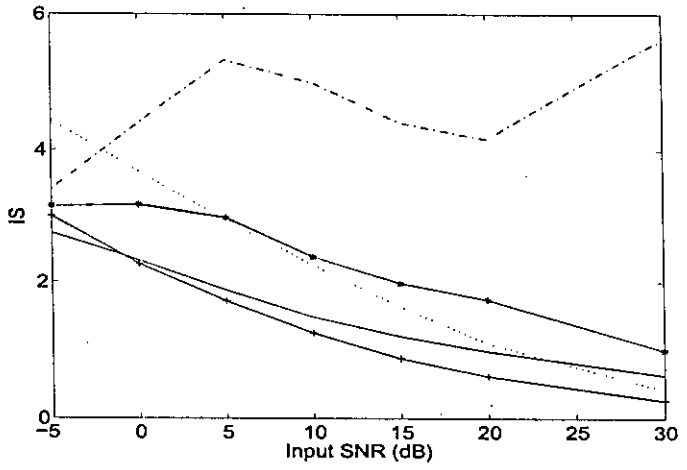


(b)

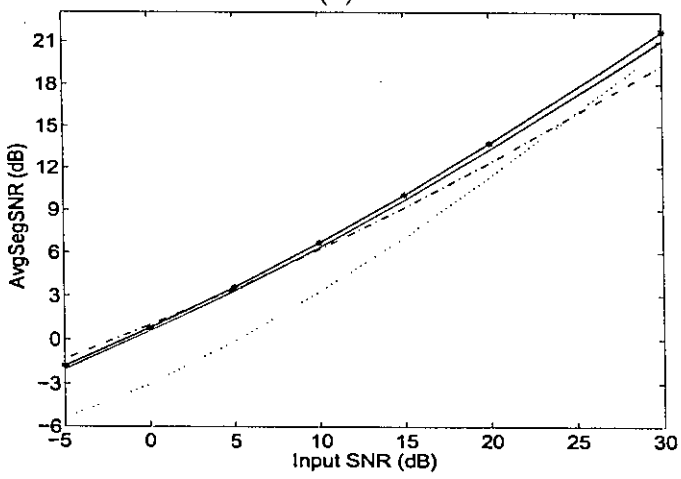


(c)

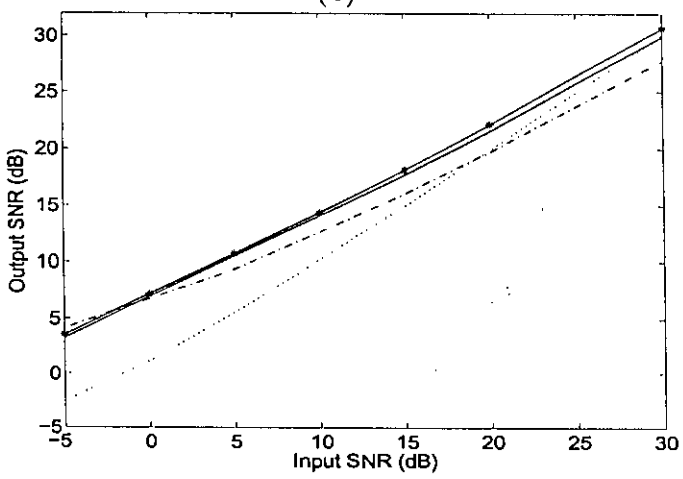
Fig. 3.13: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PARA, Wiener, dual gain Wiener where (...) for degraded, $(+-)$ for PARA using $\alpha_{n,k}$, $(*-)$ for Wiener using $\alpha_{n,k}$, $(-.)$ for dual using $\alpha = 0.98$ and $(-)$ for dual using $\alpha_{n,k}$ (S1, white noise).



(a)

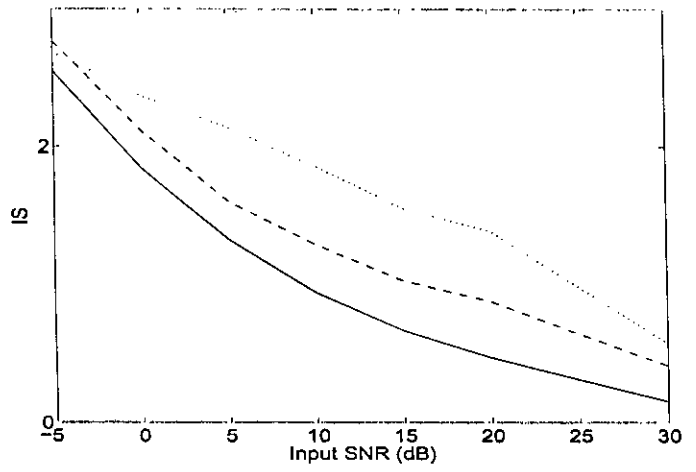


(b)

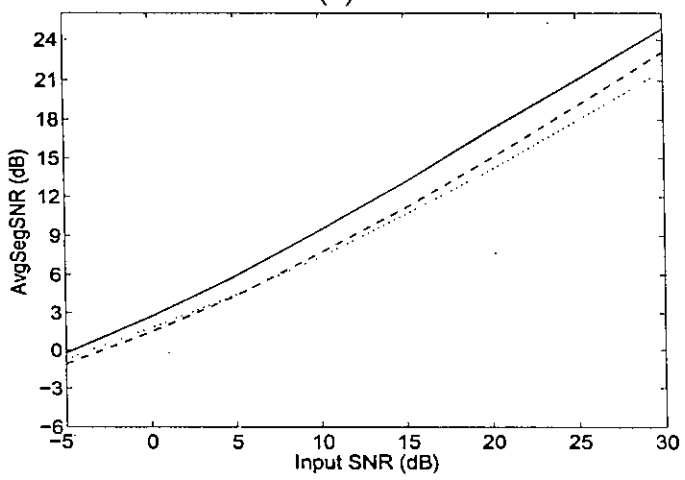


(c)

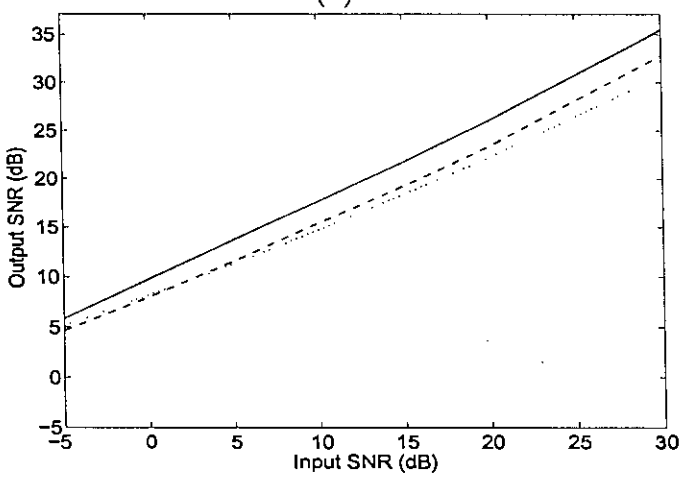
Fig. 3.14: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for PARA, Wiener, dual gain Wiener where (...) for degraded, (+−) for PARA using $\alpha_{n,k}$, (∗−) for Wiener using $\alpha_{n,k}$, (−.) for dual using $\alpha = 0.98$ and (−) for dual using $\alpha_{n,k}$ (S1, babble noise).



(a)



(b)



(c)

Fig. 3.15: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for Wiener, dual gain Wiener and constraint dual gain Wiener where (...) for Wiener using $\alpha_{n,k}$, $(--)$ for dual using $\alpha_{n,k}$ and $(-)$ for constraint dual Wiener filter using $\alpha_{n,k}$ (S1, white noise).

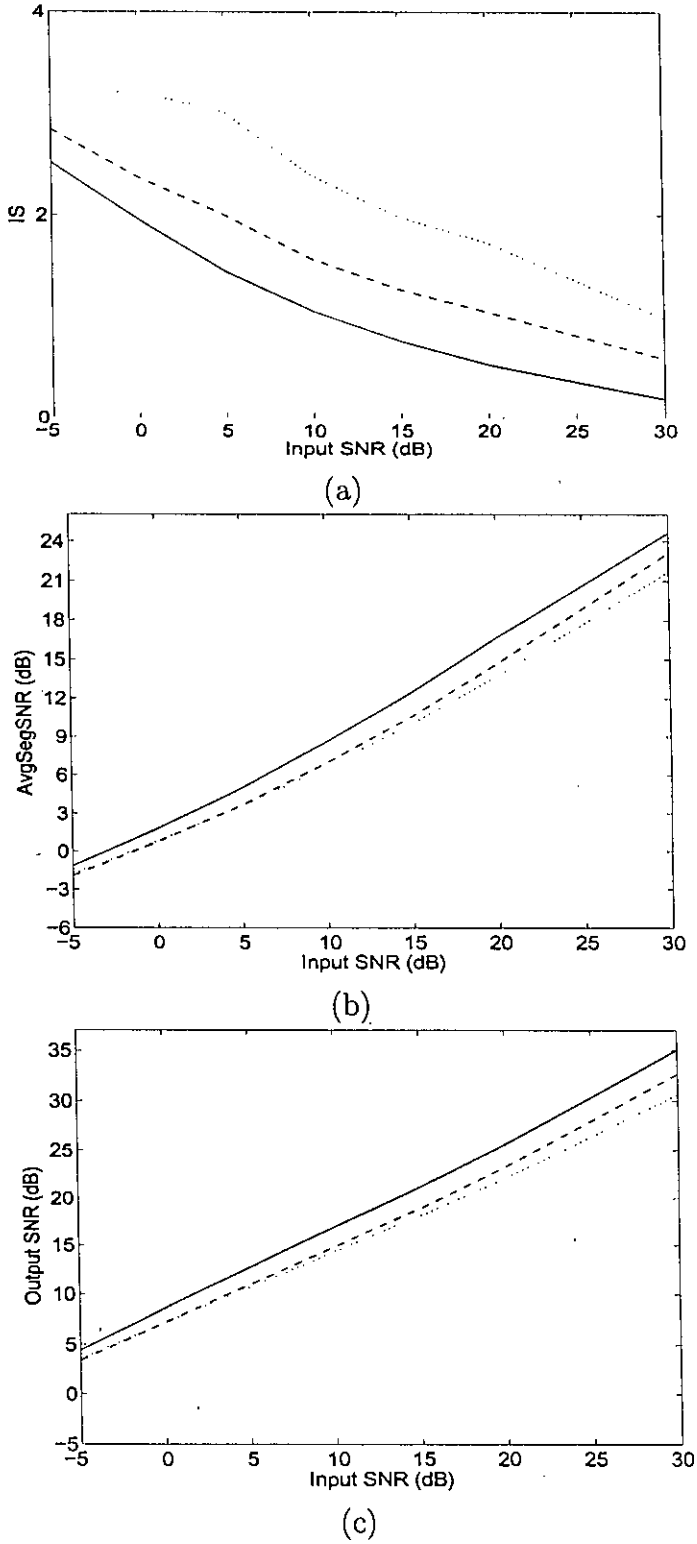
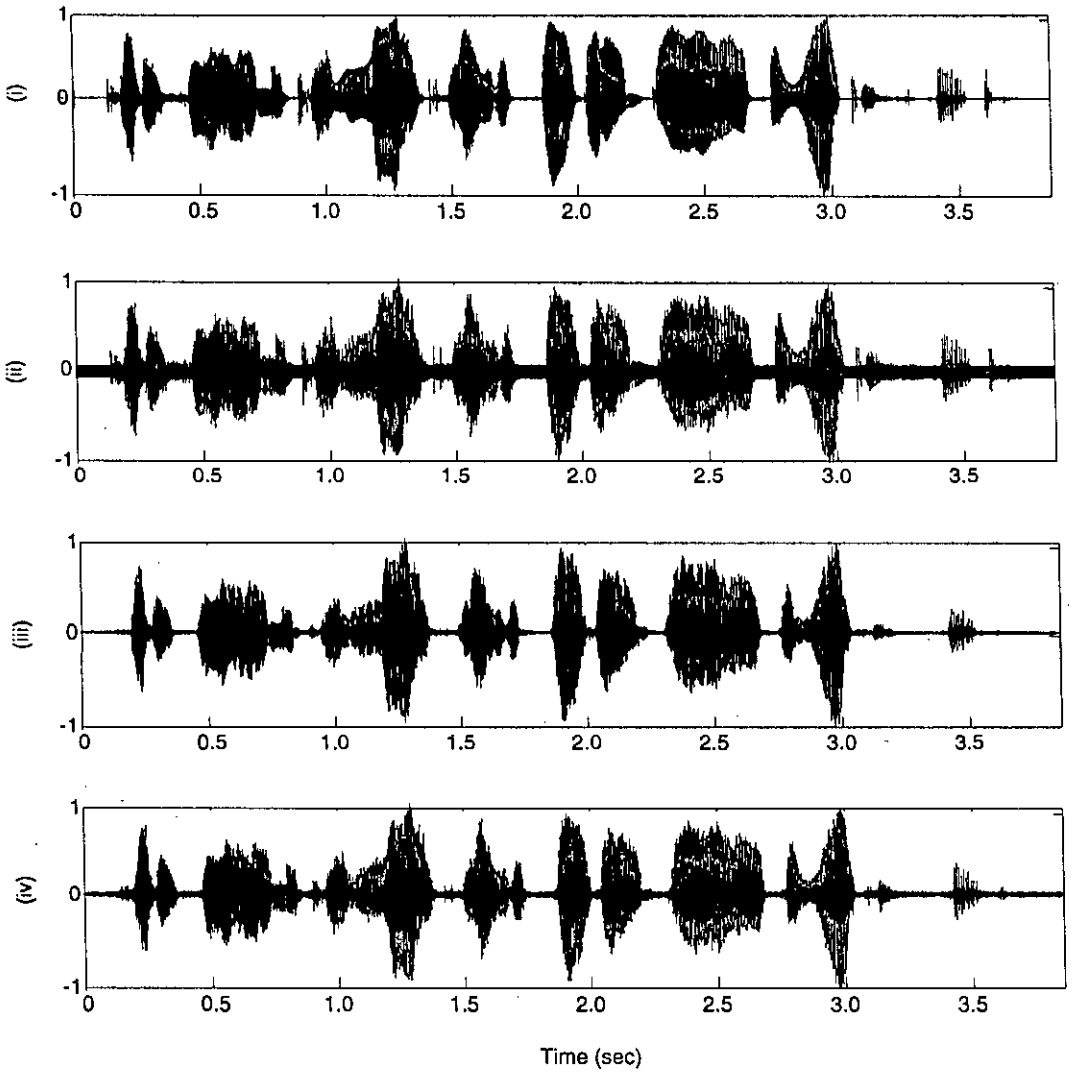
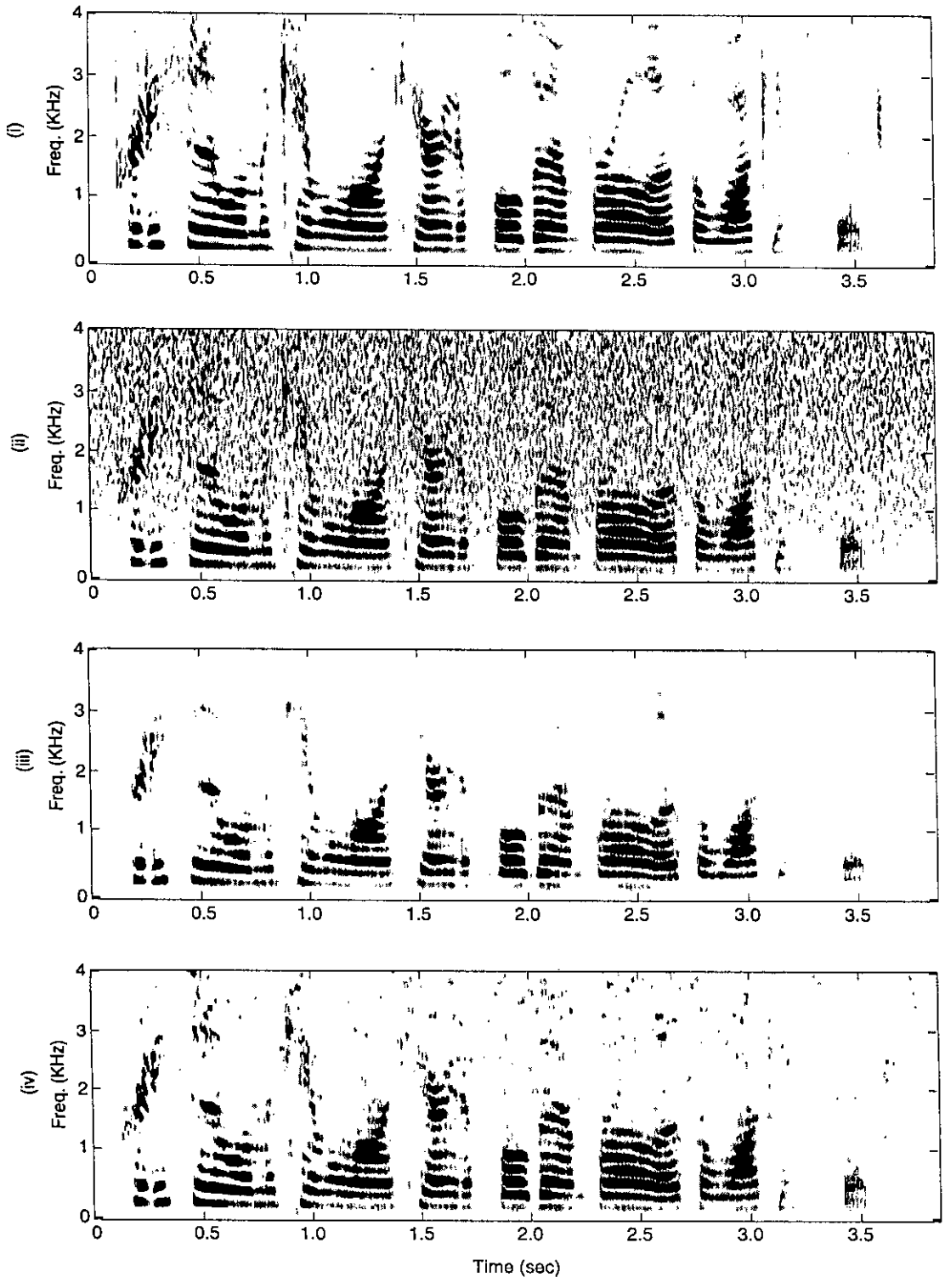


Fig. 3.16: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for Wiener, dual gain Wiener and constraint dual gain Wiener where (...) for Wiener using $\alpha_{n,k}$, (−−) for dual using $\alpha_{n,k}$ and (−) for constraint dual Wiener filter using $\alpha_{n,k}$ (S1, babble noise).



(a)



(b)

Fig. 3.17: Enhancement results for female utterance "Pretty soon a woman came along with carrying a folded umbrella as a walking stick" corrupted by white noise at SNR = 10 dB; (a) Time-domain; (b) Spectrogram; (i) clean, (ii) degraded, (iii) MPE using $\alpha = 0.98$, (iv) MPE using $\alpha_{n,k}$.

3.5 Conclusion

In this chapter, an optimal averaging parameter to estimate the *a priori* SNR has been proposed and derived in the MMSE sense. The performance of this parameter in the spectral subtraction rules (i.e., PE, MPE, PARA), Wiener filter and dual gain Wiener filter are shown in terms of necessary measures (IS, AvgSegSNR and overall SNR). It can be concluded that the time-frequency varying smoothing parameter causes a significant improvement in terms of all speech quality indices (i.e., IS, LAR, AvgSegSNR, overall SNR) over all methods using conventional smoothing parameter. It has been observed that improvements in IS measures of the PARA method and AvgSegSNRs and overall SNRs of the Wiener filter are more significant. Note that the IS measures of the Wiener filter are not optimal with compared to its AvgSegSNRs and overall SNRs. Intensive analysis of the underlying problem reveals that the basic assumption of uncorrelation between clean signal spectral component and noise spectral component should be relaxed. This leads us to generalize the Wiener filter. A generalized Wiener filter is proposed by relaxing the basic assumption in Chapter 4 and better quality of enhanced speech is expected. A comparative study with the spectral subtraction algorithms and Wiener filter is provided to demonstrate the effectiveness of the generalized Wiener filter.

Chapter 4

Generalized Wiener Filter

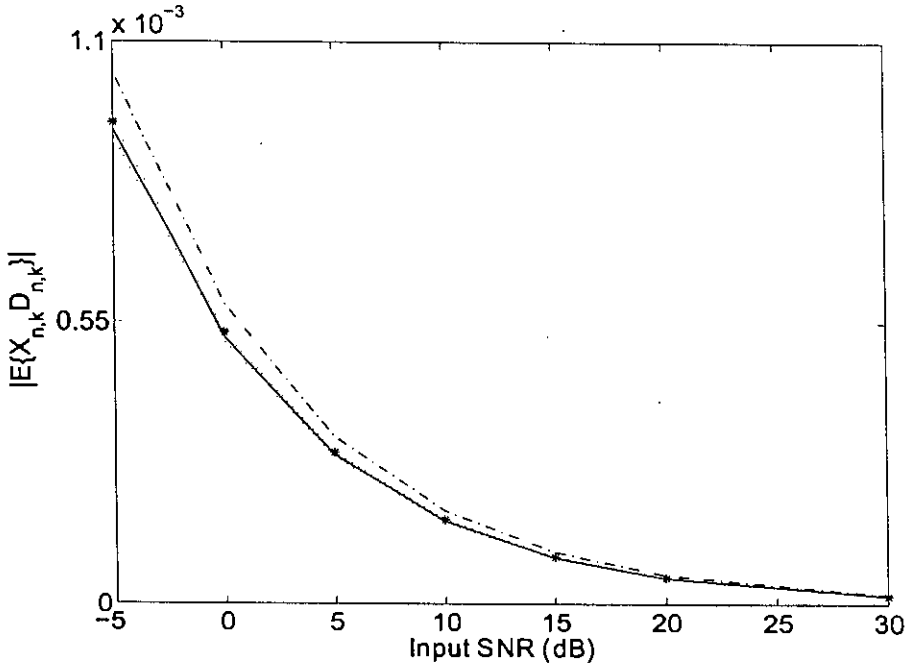
4.1 Introduction

In Chapter 3, it is shown that the single gain Wiener filter with improved estimate of the *a priori* SNR provides best overall SNR and AvgSegSNR over PE, MPE and PARA methods. But the IS measure of the enhanced speech using the Wiener filter is comparatively very high. In the derivation of the conventional Wiener filter gain it is assumed that clean and noise spectral components are uncorrelated, i.e.,

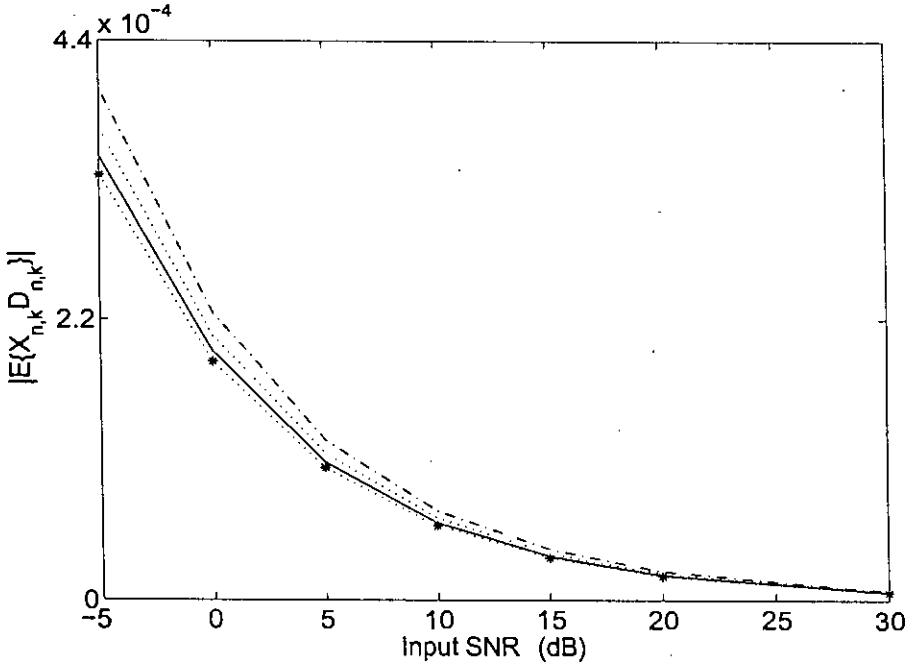
$$E\{X_{n,k}D_{n,k}\} = 0 \tag{4.1}$$

Eq. (4.1) holds when the observation noise and speech are 'truly' uncorrelated random processes. But for some color noises, namely, babble, highway, aircockpit, Eq. (4.1) holds only approximately. In particular, at low SNRs (e.g., $<$ 10 dB) it may be unrealistic to assume that the speech and noise coefficients are uncorrelated. To show this we plot $|E\{X_{n,k}D_{n,k}\}|$ at different SNRs for two speech in Fig. (4.1).

In this chapter, we take into account the correlation that exists between speech and noise at any SNR in deriving the Wiener filter gain. It is expected that the proposed generalized Wiener filter will perform better in denoising particularly at low SNRs as compared to its conventional counterpart and other spectral subtraction based methods, e.g., PE, MPE and PARA.



(a)



(b)

Fig. 4.1: Variation of $|E\{X_{n,k}D_{n,k}\}|$ with SNR; (a) S1, (b) S2 where (..) for white, (−.) for babble, (∗.) for highway and (−) for aircokcpit.

4.2 Generalized Wiener Filter Gain

Let $W_{n,k}^g$ denotes the generalized Wiener filter gain and $Y_{n,k}$ denotes noisy speech spectral component. Then an estimate of the clean speech spectral component can be obtained as

$$\widehat{X}_{n,k} = W_{n,k}^g Y_{n,k} \tag{4.2}$$

To derive $W_{n,k}^g$ in the minimum mean-square error (MMSE) sense, we minimize the cost function

$$\begin{aligned} N_{n,k} &= E\{(\widehat{X}_{n,k} - X_{n,k})^2\} \\ &= E\{(W_{n,k}^g Y_{n,k} - X_{n,k})^2\} \\ &= E\{(W_{n,k}^g (X_{n,k} + D_{n,k}) - X_{n,k})^2\} \\ &= ((W_{n,k}^g)^2 - 2W_{n,k}^g + 1)E\{X_{n,k}^2\} \\ &\quad +2W_{n,k}^g (W_{n,k}^g - 1)E\{X_{n,k}D_{n,k}\} \\ &\quad +(W_{n,k}^g)^2 E\{D_{n,k}^2\} \end{aligned} \tag{4.3}$$

Differentiating $N_{n,k}$ with respect to $W_{n,k}^g$ gives

$$\begin{aligned} \frac{\partial N_{n,k}}{\partial W_{n,k}^g} &= (2W_{n,k}^g - 2)E\{X_{n,k}^2\} + 2(2W_{n,k}^g - 1)E\{X_{n,k}D_{n,k}\} \\ &\quad +2W_{n,k}^g E\{D_{n,k}^2\} \end{aligned} \tag{4.4}$$

Now equating $\partial N_{n,k}/\partial W_{n,k}^g$ to zero yields

$$2(W_{n,k}^g - 1)E\{X_{n,k}^2\} + 2(2W_{n,k}^g - 1)E\{X_{n,k}D_{n,k}\} + 2W_{n,k}^g E\{D_{n,k}^2\} = 0 \tag{4.5}$$

Dividing Eq. (4.5) by $2E\{D_{n,k}^2\}$, we obtain

$$2(W_{n,k}^g - 1)\frac{E\{X_{n,k}^2\}}{2E\{D_{n,k}^2\}} + \frac{2(2W_{n,k}^g - 1)E\{X_{n,k}D_{n,k}\}}{2E\{D_{n,k}^2\}} + W_{n,k}^g = 0 \tag{4.6}$$

Substituting $E\{X_{n,k}^2\}/E\{D_{n,k}^2\} = \xi_{n,k}$ (defined in Eq. (2.13)) into Eq. (4.6) gives

$$W_{n,k}^g \left\{ \xi_{n,k} + 1 + 2\frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}} \right\} - \xi_{n,k} - \frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}} = 0 \tag{4.7}$$

Finally, an expression for $W_{n,k}^g$ assuming $E\{X_{n,k}D_{n,k}\} \neq 0$ is obtained as

$$W_{n,k}^g = \frac{\xi_{n,k} + \frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}}}{\xi_{n,k} + 1 + 2\frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}}} \tag{4.8}$$

In the following, we propose a recursive technique to estimate $E\{X_{n,k}D_{n,k}\}$. We can write

$$\begin{aligned} E\{Y_{n,k}D_{n,k}\} &= E\{(X_{n,k}+D_{n,k})D_{n,k}\} \\ &= E\{X_{n,k}D_{n,k}\}+E\{D_{n,k}^2\} \end{aligned} \tag{4.9}$$

Dividing Eq. (4.9) by $E\{D_{n,k}^2\}$, we obtain

$$\frac{E\{Y_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}} = \frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}}+1 \tag{4.10}$$

Rearranging

$$\begin{aligned} \frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}} &= \frac{E\{Y_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}}-1 \\ &= T_{n,k}-1 \end{aligned} \tag{4.11}$$

where

$$\begin{aligned} T_{n,k} &= \frac{E\{Y_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}} \\ &= \frac{E\{Y_{n,k}(Y_{n,k}-X_{n,k})\}}{E\{D_{n,k}^2\}} \\ &= \frac{E\{Y_{n,k}^2\}}{E\{D_{n,k}^2\}}-\frac{E\{Y_{n,k}X_{n,k}\}}{E\{D_{n,k}^2\}} \end{aligned} \tag{4.12}$$

In this work, we estimate $T_{n,k}$ recursively as

$$\begin{aligned} \widehat{T}_{n,k} &= \beta_{n,k}\widetilde{T}_{n-1,k}+(1-\beta_{n,k})\left[\frac{Y_{n,k}^2}{E\{D_{n,k}^2\}}-\frac{Y_{n,k}X_{n,k}}{E\{D_{n,k}^2\}}\right] \\ &= \beta_{n,k}\widetilde{T}_{n-1,k}+(1-\beta_{n,k})\left[\gamma_{n,k}-\frac{X_{n,k}}{Y_{n,k}}\frac{Y_{n,k}^2}{E\{D_{n,k}^2\}}\right] \\ &= \beta_{n,k}\widetilde{T}_{n-1,k}+(1-\beta_{n,k})\left[\gamma_{n,k}-\gamma_{n,k}\frac{X_{n,k}}{Y_{n,k}}\right] \\ &= \beta_{n,k}\widetilde{T}_{n-1,k}+(1-\beta_{n,k})\gamma_{n,k}\left[1-\frac{X_{n,k}}{Y_{n,k}}\right] \end{aligned} \tag{4.13}$$

where $\widetilde{T}_{n-1,k}$ is the previous frame estimate, $\beta_{n,k}$ is the averaging parameter, $0 \le \beta_{n,k} \le 1$. Substituting Eq. (4.11) into Eq. (4.8), we obtain an optimum expression for the gain function of the generalized Wiener filter

$$\begin{aligned} W_{n,k}^{g} &= \frac{\xi_{n,k}+T_{n,k}-1}{\xi_{n,k}+1+2T_{n,k}-2} \\ &= \frac{\xi_{n,k}-1+T_{n,k}}{\xi_{n,k}-1+2T_{n,k}} \end{aligned} \tag{4.14}$$

Notice that the *a priori* SNR $\xi_{n,k}$ is estimated using the "decision-directed" approach as described in Eq. (3.5).

In the following section, an optimum expression for the time-frequency varying averaging parameter $\beta_{n,k}$ in the MMSE sense (in the DCT domain) is derived.

4.3 Estimating $\beta_{n,k}$

It is desired that the estimate $\widehat{T}^p_{n,k}$ should actually be as close as possible to $T_{n,k}$. Here we propose an MMSE estimator for $\beta_{n,k}$ which minimizes the cost function

$$\begin{aligned} J_\beta &= E\left\{(\widehat{T}^p_{n,k} - T_{n,k})^2 \mid \widetilde{T}_{n-1,k}\right\} \\ &= E\left\{(\widehat{T}^p_{n,k})^2 - 2\widehat{T}^p_{n,k}T_{n,k} + T^2_{n,k}) \mid \widetilde{T}_{n-1,k}\right\} \end{aligned} \tag{4.15}$$

given $\widetilde{T}_{n-1,k}$. Substituting Eq. (4.13) into Eq. (4.15), we obtain

$$\begin{aligned} J_\beta &= E\left\{\left(\beta^2_{n,k}\widetilde{T}^2_{n-1,k} + (1-\beta_{n,k})^2\gamma^2_{n,k}\chi^2_{n,k} + 2\beta_{n,k}(1-\beta_{n,k})\widetilde{T}_{n-1,k}\gamma_{n,k}\chi_{n,k}\right.\right. \\ &\quad \left.\left. -2T_{n,k}[\beta_{n,k}\widetilde{T}_{n-1,k} + (1-\beta_{n,k})\gamma_{n,k}\chi_{n,k}] + T^2_{n,k}\right) \mid \widetilde{T}_{n-1,k}\right\} \\ &= \beta^2_{n,k}\widetilde{T}^2_{n-1,k} + (1-\beta_{n,k})^2 E\left\{\gamma^2_{n,k}\right\}\chi^2_{n,k} + 2\beta_{n,k}(1-\beta_{n,k})\widetilde{T}_{n-1,k}E\left\{\gamma_{n,k}\right\}\chi_{n,k} \\ &\quad -2\beta_{n,k}T_{n,k}\widetilde{T}_{n-1,k} - 2(1-\beta_{n,k})T_{n,k}E\left\{\gamma_{n,k}\right\}\chi_{n,k} + T^2_{n,k} \end{aligned} \tag{4.16}$$

where $\chi_{n,k} = 1 - X_{n,k}/Y_{n,k}$, independent of $\beta_{n,k}$. Using Eq. (2.12), we can write

$$E\{\gamma^2_{n,k}\} = \frac{E\{Y^4_{n,k}\}}{E\{D^2_{n,k}\}^2} \tag{4.17}$$

and

$$E\{\gamma_{n,k}\} = \frac{E\{Y^2_{n,k}\}}{E\{D^2_{n,k}\}} \tag{4.18}$$

To find the value of $E\{\gamma^2_{n,k}\}$ we need to evaluate $E\{Y^4_{n,k}\}$. As in Chapter 3, we drop subscript (n,k) for notational simplicity. Using Eqs. (3.14) and (3.15)

$$\begin{aligned} E\{Y^4\} &= E\{X^4 + 4X^2D^2 + D^4 + 2X^3D + 2XD^3 + 2X^2D^2\} \\ &= E\{X^4\} + 4E\{X^2\}E\{D^2\} + E\{D^4\} + 2E\{X^3\}E\{D\} \\ &\quad +2E\{X\}E\{D^3\} + 2E\{X^2\}E\{D^2\} \end{aligned} \tag{4.19}$$

Using the fact that $X_{n,k}$ and $D_{n,k}$ are zero-mean but uncorrelated real Gaussian random variables, the simplified form of $E\{Y^4\}$ is obtained as

$$\begin{aligned} E\{Y^4\} &= E\{X^4\} + 4E\{X^2\}E\{D^2\} + E\{D^4\} + 2E\{X^2\}E\{D^2\} \\ &= E\{X^4\} + E\{D^4\} + 6E\{X^2\}E\{D^2\} \end{aligned} \tag{4.20}$$

We know from Eqs. (3.18) and (3.19)

$$E\{X_{n,k}^4\} = 3E\{X_{n,k}^2\}^2 \tag{4.21}$$

and

$$E\{D_{n,k}^4\} = 3E\{D_{n,k}^2\}^2 \tag{4.22}$$

Substituting Eqs. (4.21) and (4.22) into Eq. (4.20) yields

$$E\{Y_{n,k}^4\} = 3E\{X_{n,k}^2\}^2 + 3E\{D_{n,k}^2\}^2 + 6E\{X_{n,k}^2\}E\{D_{n,k}^2\} \tag{4.23}$$

Substituting Eq. (4.23) into Eq. (4.17), we get

$$\begin{aligned} E\{\gamma_{n,k}^2\} &= \frac{3E\{X_{n,k}^2\} + 3E\{D_{n,k}^2\} + 6E\{X_{n,k}^2\}E\{D_{n,k}^2\}}{E\{D_{n,k}^2\}^2} \\ &= 3\frac{E\{X_{n,k}^2\}^2}{E\{D_{n,k}^2\}^2} + 3\frac{E\{D_{n,k}^2\}^2}{E\{D_{n,k}^2\}^2} + 6\frac{E\{X_{n,k}^2\}E\{D_{n,k}^2\}}{E\{D_{n,k}^2\}^2} \\ &= 3(\frac{E\{X_{n,k}^2\}}{E\{D_{n,k}^2\}})^2 + 3\frac{E\{D_{n,k}^2\}^2}{E\{D_{n,k}^2\}} + 6\frac{E\{X_{n,k}^2\}}{E\{D_{n,k}^2\}} \\ &= 3\xi_{n,k}^2 + 3 + 6\xi_{n,k} \\ &= 3(\xi_{n,k}+1)^2 \end{aligned} \tag{4.24}$$

Substituting Eq. (3.14) into Eq. (4.18), we obtain

$$\begin{aligned} E\{\gamma_{n,k}\} &= \frac{E\{Y_{n,k}^2\}}{E\{D_{n,k}^2\}} \\ &= \frac{E\{(X_{n,k}^2 + 2X_{n,k}D_{n,k} + D_{n,k}^2)\}}{E\{D_{n,k}^2\}} \\ &= \frac{E\{X_{n,k}^2\}}{E\{D_{n,k}^2\}} + 2\frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}} + \frac{E\{D_{n,k}^2\}}{E\{D_{n,k}^2\}} \\ &= \xi_{n,k} + 2\frac{E\{X_{n,k}D_{n,k}\}}{E\{D_{n,k}^2\}} + 1 \end{aligned} \tag{4.25}$$

Combining Eqs. (4.25) and (4.11)

$$\begin{aligned} E\{\gamma_{n,k}\} &= \xi_{n,k} + 2T_{n,k} - 2 + 1 \\ &= \xi_{n,k} + 2T_{n,k} - 1 \end{aligned} \tag{4.26}$$

Now substituting $E\{\gamma_{n,k}\}$ and $E\{\gamma_{n,k}^2\}$ into Eq. (4.16), we obtain

$$\begin{aligned} J_\beta &= \beta_{n,k}^2\tilde{T}_{n-1,k}^2 + (1-\beta_{n,k})^2 3(\xi_{n,k}+1)^2\chi_{n,k}^2 \\ &\quad +2\beta_{n,k}(1-\beta_{n,k})\tilde{T}_{n-1,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k} - 2\beta_{n,k}T_{n,k}\tilde{T}_{n-1,k} \\ &\quad -2(1-\beta_{n,k})T_{n,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k} + T_{n,k}^2 \end{aligned} \tag{4.27}$$

Differentiating J_β with respect to $\beta_{n,k}$

$$\begin{aligned}\frac{\partial J_\beta}{\partial \beta_{n,k}} = \quad & 2\beta_{n,k}\tilde{T}^2_{n-1,k} - 2(1-\beta_{n,k})3(\xi_{n,k}+1)^2\chi^2_{n,k} \\ & +2(1-2\beta_{n,k})\tilde{T}_{n-1,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k} \\ & -2T_{n,k}\tilde{T}_{n-1,k} + 2T_{n,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k} \\ = \quad & 2\beta_{n,k}\left[\tilde{T}^2_{n-1,k} + 3(\xi_{n,k}+1)^2\chi^2_{n,k}\right. \\ & \left.-2\tilde{T}_{n-1,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k}\right] - 2\left[3(\xi_{n,k}+1)^2\chi^2_{n,k}\right. \\ & -\tilde{T}_{n-1,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k} + T_{n,k}\tilde{T}_{n-1,k} \\ & \left.-T_{n,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k}\right] \end{aligned} \tag{4.28}$$

Now equating $\partial J_\beta/\partial\beta_{n,k}$ to zero, the optimum expression of $\beta_{n,k}$ is obtained as

$$\beta_{n,k} = \frac{3(\xi_{n,k}+1)^2\chi^2_{n,k} + T_{n,k}\tilde{T}_{n-1,k} - (T_{n,k}+\tilde{T}_{n-1,k})(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k}}{\tilde{T}^2_{n-1,k} + 3(\xi_{n,k}+1)^2\chi^2_{n,k} - 2\tilde{T}_{n-1,k}(\xi_{n,k}+2T_{n,k}-1)\chi_{n,k}} \tag{4.29}$$

4.3.1 Implementation of $\beta_{n,k}$

In the above expression $\chi_{n,k} = 1 - X_{n,k}/Y_{n,k}$ and $T_{n,k}$ are unknown. $\chi_{n,k}$ is not computable as $X_{n,k}$ is unknown. However $\chi_{n,k}$ may be replaced by $1 - W_{n,k}$, where $W_{n,k}$ is Wiener gain calculated using our proposed smoothing parameter to estimate the *a priori* SNR. Hence $T_{n,k}$ may be replaced by $\gamma_{n,k}\chi_{n,k}$ (comparing Eqs. (4.12) and (4.13)).

4.4 Simulation Results and Discussions

In this result section, we show simulation results for the same speech and noises as in the previous chapter.

Fig. (4.2) shows the variation of $\beta_{n,k}$ for all the speech frames. It can be observed that smoothing parameter $\beta_{n,k}$ lies in the range of $0 \leq \beta_{n,k} \leq 1$ as expected. We present comparative results of the proposed generalized Wiener filter with the PE, MPE, PARA and conventional Wiener filter methods in Figs. (4.3)-(4.6). It may be restated that the main purpose of proposing the generalized Wiener filter is to improve the IS measure of the conventional Wiener filter without sacrificing output SNR and AvgSegSNR. As can be seen from the figures, the IS index has significantly improved for both the utterances and the noises used,

and at all SNRs (-5 dB to 30 dB). AvgSegSNR and overall output SNR are still better for the generalized Wiener filter than the PE, MPE, PARA methods, but comparable with the Wiener filter as desired.

Finally, we present speech enhancement results in the time and frequency domain. Fig. 4.7 (a) shows the highway noise degraded speech $y(t)$ at SNR = 10 dB for the female utterance ("Pretty soon a woman came along with a folded umbrella as a walking stick"), and the enhanced speech resulting from the proposed generalized Wiener filter and the conventional Wiener filter. Fig. 4.7 (b) shows the corresponding spectrograms. As expected, proposed generalized filter produces lower residual noise and noticeably less speech distortion in some speech segments.

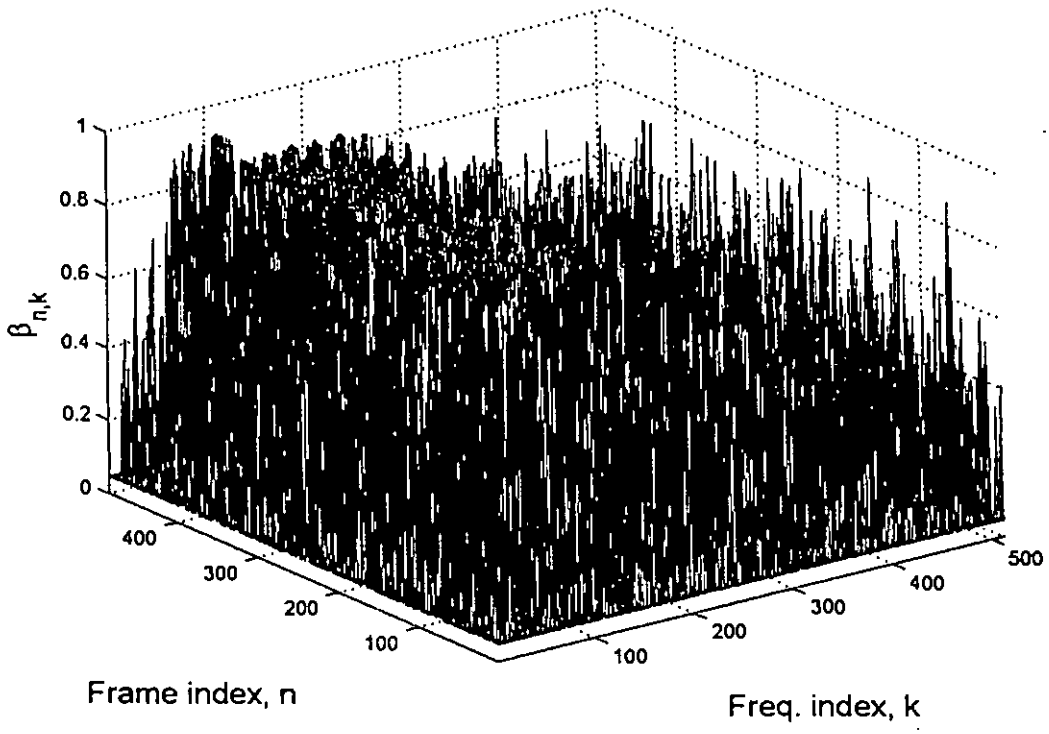
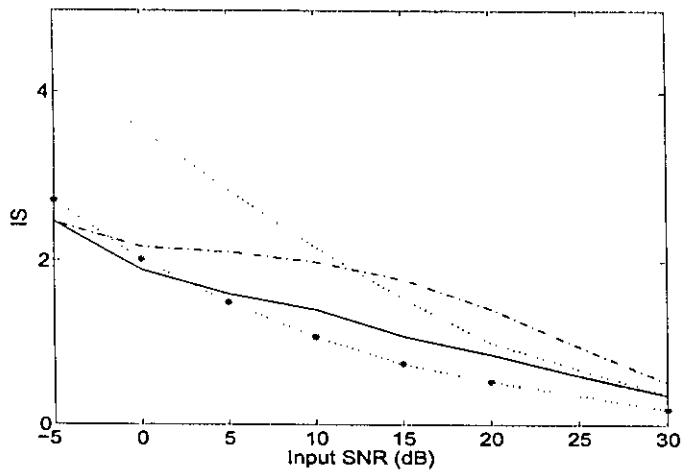
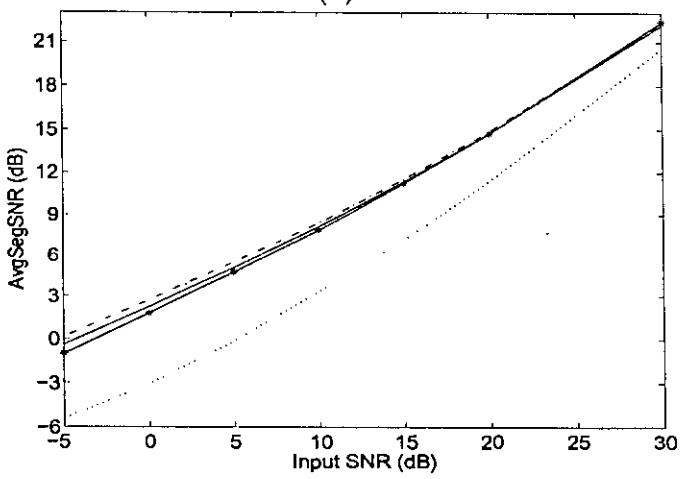


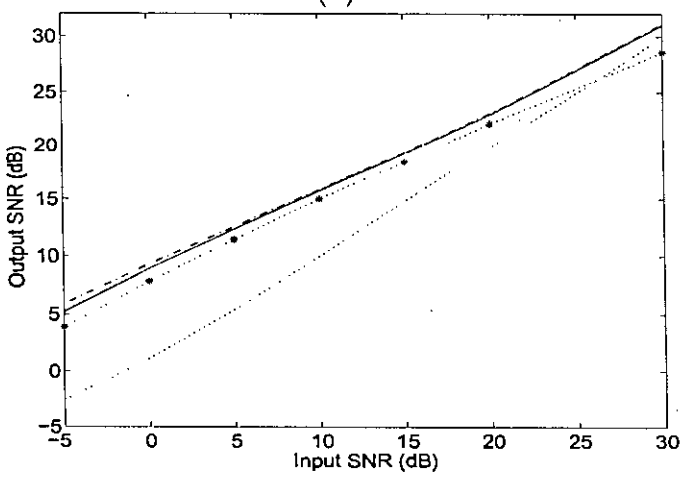
Fig. 4.2: Variation of $\beta_{n,k}$ for S1 corrupted by white noise at SNR=10 dB of the generalized Wiener filter.



(a)

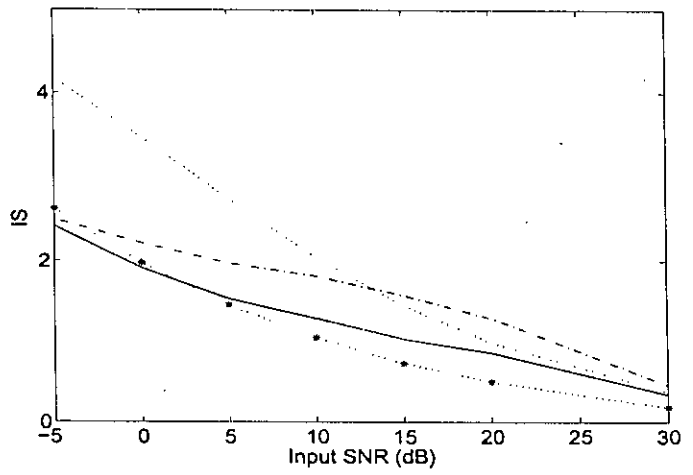


(b)

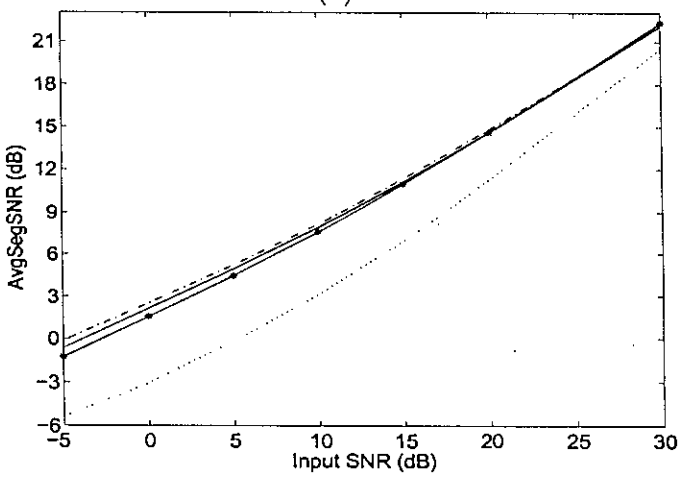


(c)

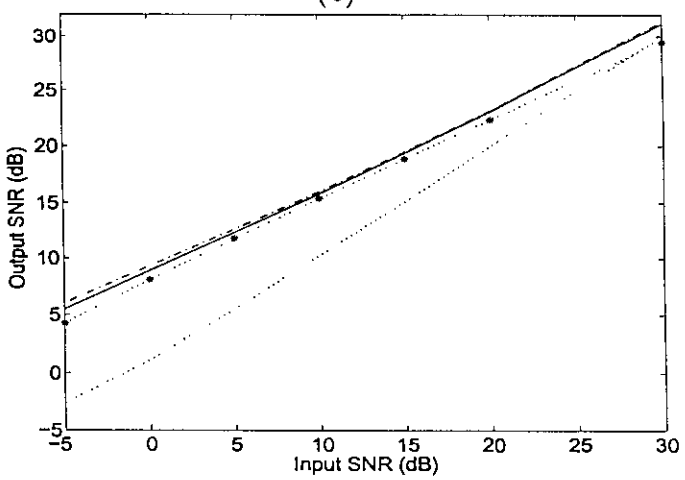
Fig. 4.3: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA, Wiener and generalized Wiener where (...) for degraded, $(*.)$ for PARA using $\alpha_{n,k}$, $(*-)$ for MPE using $\alpha_{n,k}$, $(-.)$ for Wiener filter using $\alpha_{n,k}$ and $(-)$ for generalized Wiener filter using $\alpha_{n,k}$ (S1, highway noise).



(a)

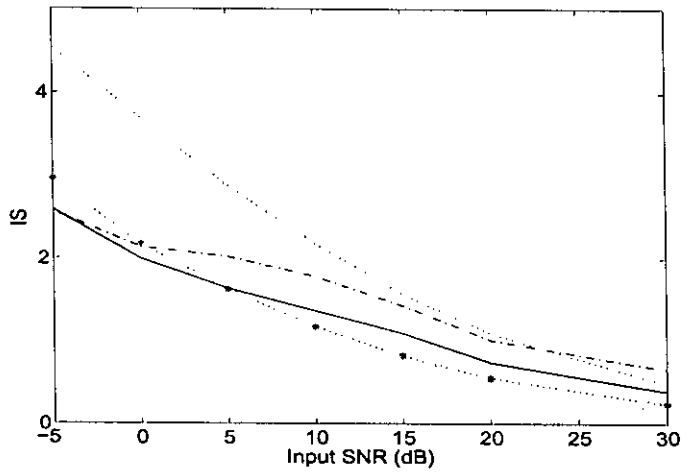


(b)

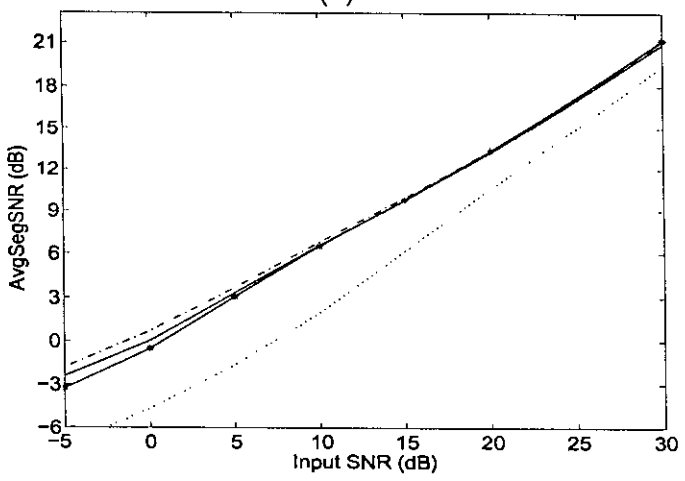


(c)

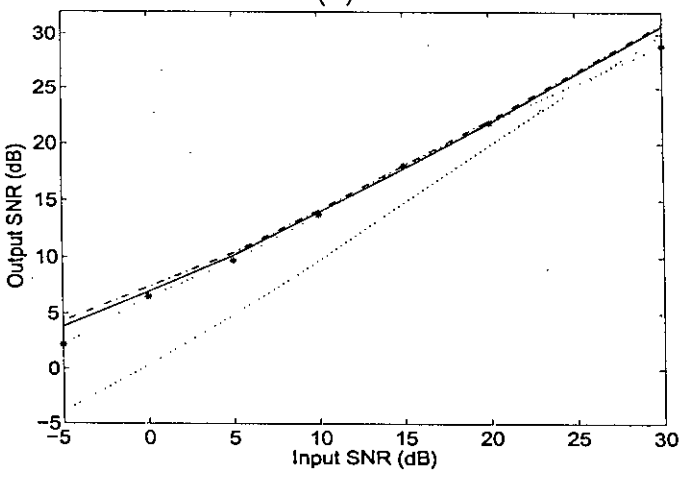
Fig. 4.4: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA, Wiener and generalized Wiener where (...) for degraded, (*.) for PARA using $\alpha_{n,k}$, (*−) for MPE using $\alpha_{n,k}$, (−.) for Wiener filter using $\alpha_{n,k}$ and (−) for generalized Wiener filter using $\alpha_{n,k}$ (S1, aircockpit noise).



(a)

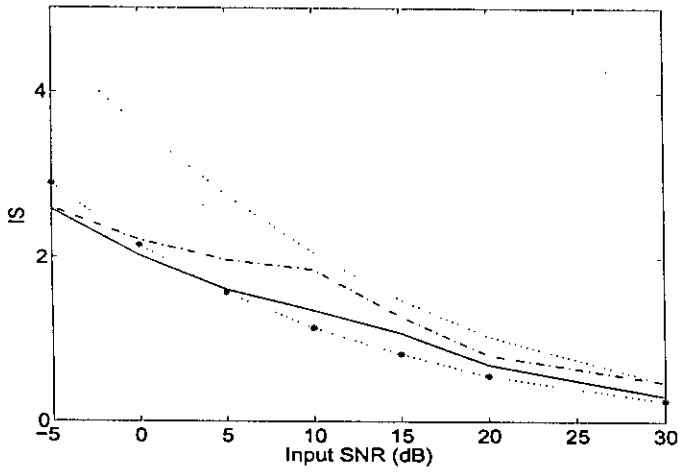


(b)

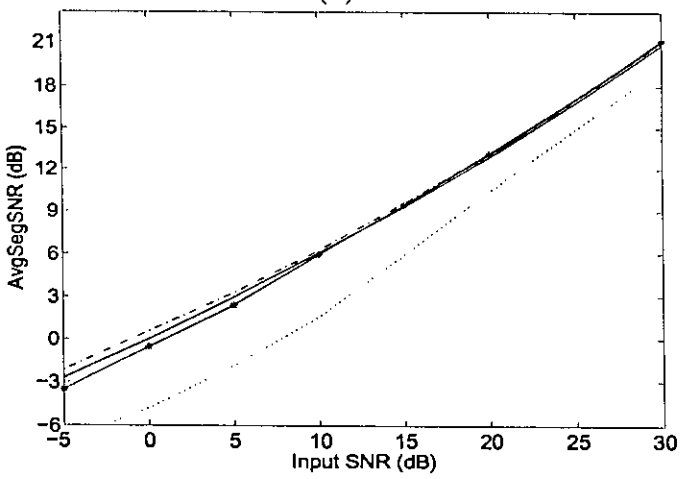


(c)

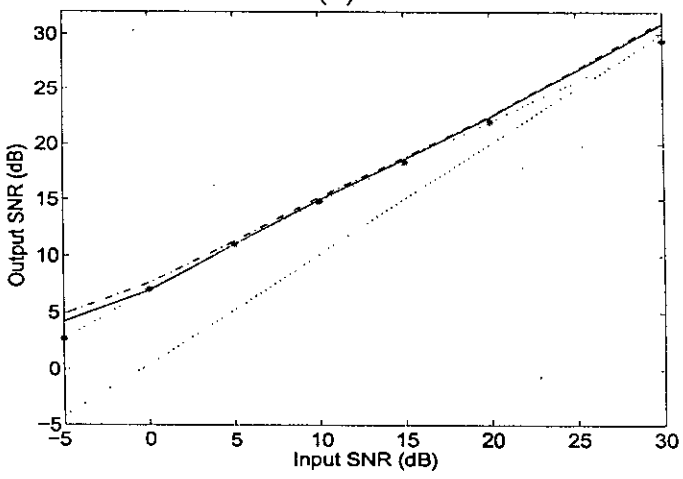
Fig. 4.5: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA, Wiener and generalized Wiener where (...) for degraded, $(*.)$ for PARA using $\alpha_{n,k}$, $(*-)$ for MPE using $\alpha_{n,k}$, $(-.)$ for Wiener filter using $\alpha_{n,k}$ and $(-)$ for generalized Wiener filter using $\alpha_{n,k}$ (S2, highway noise).



(a)

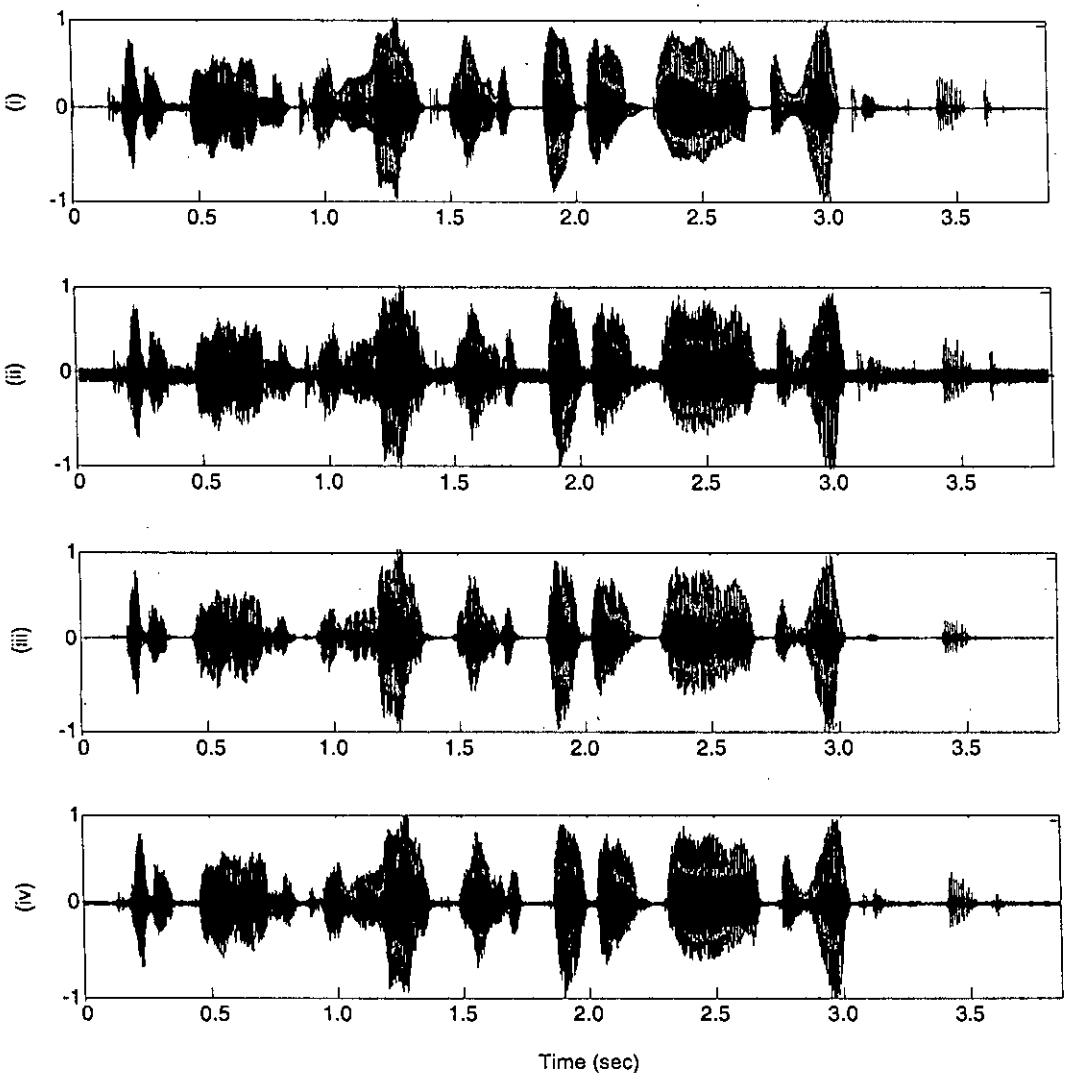


(b)

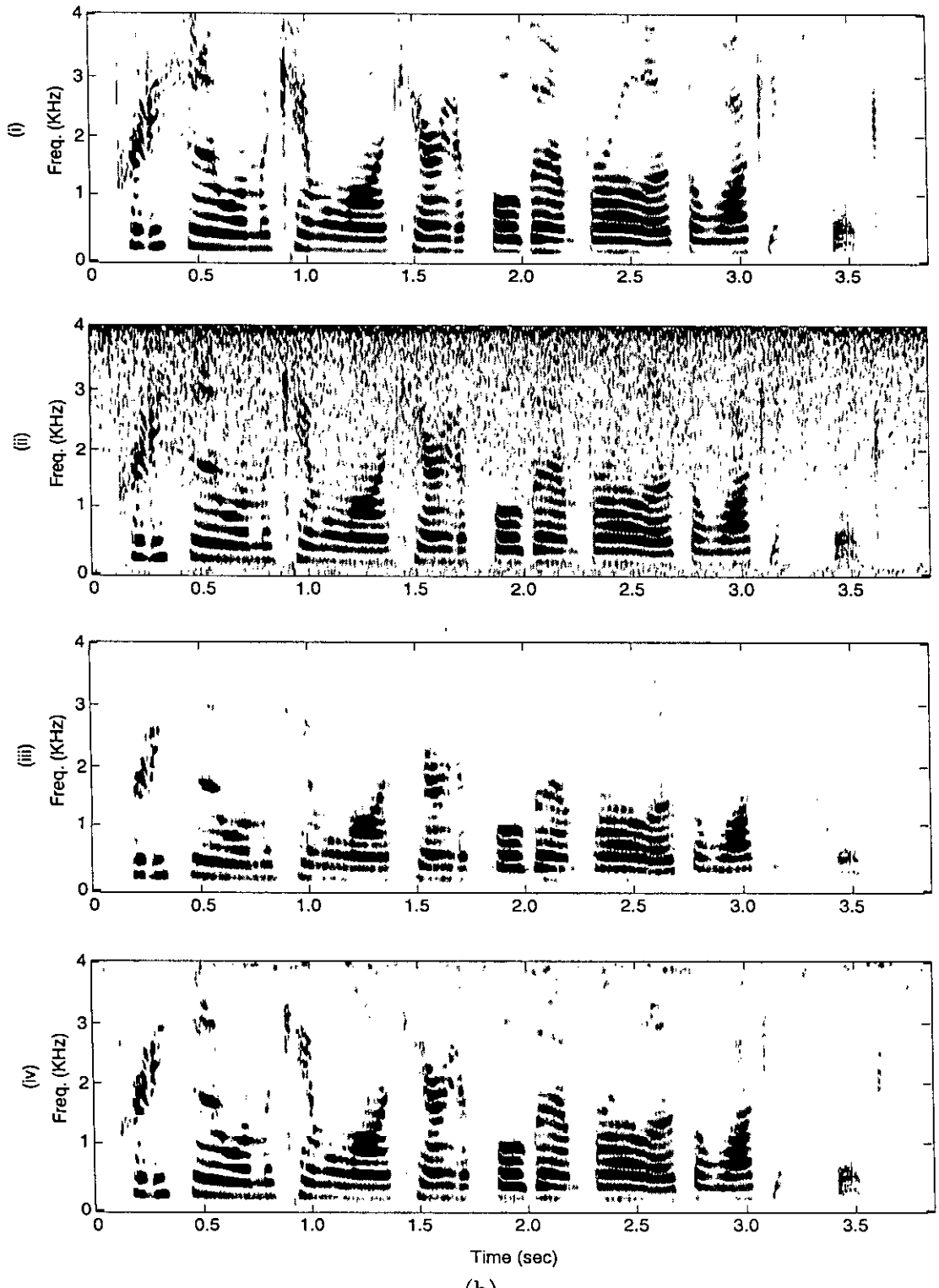


(c)

Fig. 4.6: Variation of (a) IS, (b) AvgSegSNR, (c) Output SNR for MPE, PARA, Wiener and generalized Wiener where (...) for degraded, (*.) for PARA using $\alpha_{n,k}$, (*−) for MPE using $\alpha_{n,k}$, (−.) for Wiener filter using $\alpha_{n,k}$ and (−) for generalized Wiener filter using $\alpha_{n,k}$ (S2, aircockpit noise).



(a)



(b)

Fig. 4.7: Enhancement results for female utterance "Pretty soon a woman came along with carrying a folded umbrella as a walking stick" corrupted by highway noise at SNR = 10 dB; (a) Time-domain; (b) Spectrogram; (i) clean, (ii) degraded, (iii) Wiener using $\alpha = 0.98$, (iv) generalized Wiener.

4.5 Conclusion

In this Chapter, we have presented a generalized Wiener filter for improving the IS index of the conventional Wiener filter. Simulation results shown for different speech and noises reveal that the proposed generalization to the conventional Wiener filter is particularly useful at a low SNRs.

Chapter 5

Conclusion

5.1 Summary

The major focus of this research has been to further improve the performances of the traditional spectral subtraction methods with a better estimate of the *a priori* SNR. The performances of the spectral subtraction methods, the Wiener filter and the dual gain Wiener filter incorporating optimal averaging parameter to estimate the *a priori* SNR have been reported with necessary evaluations. Significant improvement in terms of enhanced speech quality indices (i.e., IS, AvgSegSNR, Overall SNR) has been observed over all methods using conventional smoothing parameter. It has been also noticed that using our proposed smoothing parameter, the spectral subtraction methods prevent the undesired fall of SNR of the denoised speech even when the original signal has a SNR of 30 dB. It has been observed that improvements in IS measures of the PARA method and AvgSegSNRs and overall SNRs of the Wiener filter are more significant and noticeable. It has also been observed that the IS measures of the Wiener filter are not optimal as compared to its AvgSegSNRs and overall SNRs. To improvethe IS measure, a generalized Wiener filter has been proposed by relaxing the assumption that clean speech and noise spectral components are uncorrelated. The performance of the generalized Wiener filter has been evaluated using the standard TIMIT and NOISEX databases. It has been shown using several numerical examples that the generalized Wiener filter has performed optimally (i.e., in terms of all speech quality indices (e.g., IS, AvgSegSNR and overall output SNR)) compared to its conventional counterpart and other spectral subtraction based methods, e.g., PE, MPE and PARA. The improvement in terms of quality indices of the

proposed scheme has been found to be particularly significant at low SNRS.

5.2 Future Works

It has been observed that the dual gain Wiener filtering give highly impressive results in terms of quality indices (e.g., IS, AvgSegSNR, output SNR) if we can accurately identify noisy spectral components whose amplitude has been increased or decreased by noise. The prediction based algorithm suggested in [43] has been found to be ineffective in discriminating the spectral components. An effective algorithm for this purpose is highly desired.

Bibliography

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 27, pp. 113-120, 1979.

[2] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586-1604, 1979.

[3] J. Makhoul *et al., Removal of noise from noise-degraded speech signals*, Panel on removal of noise from a speech/noise signal, National Research council. Washington, DC : National academy press, 1989.

[4] D. O'shaughnessy, "Enhancing speech degraded by additive noise or interferring speakers," *IEEE Commun. Mag.*, pp. 46-52, Feb. 1989.

[5] S. F. Boll, "Speech enhancement in the 1980's: Noise suppression with pattern matching," in *Advances in Speech Signal Processing*, S. Furui and M. M. Sondhi, Eds. New York : Marcel DEkker, 1992.

[6] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proc. of IEEE*, vol. 80, no. 10, pp. 1526-1555, 1992.

[7] I. Lecomte, M. Lever, J. Boudy and A. Tassy, "Car noise processing for speech input," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.512-515, May 1989.

[8] N. D. Degan and C. Prati, "Performance of speech enhancement techniques for mobile radio terminal application," *Signal Processing III: Theories and applications*, New York: Elsevier Pulishers B. V. (North Holland), pp.381-385, 1986.

[9] R. J. Niederjohn and J. H. Grotelueschen, "Speech intelligilibility enhancement in a power generating noise environment," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 26, pp. 208-210, Aug. 1978.

[10] I. Pollack, "Speech communications at high noise levels: The role of a noise operated automatic gain control system and hearing protection," *J. Acoust. Soc. Amer.*, vol. 29, pp. 1324-1327, Dec. 1957.

[11] I. B. Thomas and W. J. Ohley, "Intelligibility enhancement through spectral weighting," *Proc. IEEE Conf. Speech, Commun. and Processing*, pp. 360-363, 1972.

[12] J. S. Lim and A. V. Oppenheim, "Reduction of quantization noise in PCM speech encoding," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28pp. 107-110, Feb 1980.

[13] Y. Ephraim and D. Malah, "Combined enhancement and adaptive transform coding of noisy speech," *Proc. Inst. Elec Engg.*, vol. 133, pt. F, no. 1, pp. 81-86, 1986.

[14] O. Cappe, "Estimation of the musical noise phenomena with the Ephraim and Malah noise suppressor," *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 345-349, 1994.

[15] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator," *IEEE Trans. Speech Audio Processing*, vol. ASSP-32, pp. 1109-1121, 1984.

[16] R. Hoeldrich and M. Lorber, "Broadband noise reduction based on spectral subtraction," *Proc. ICSPAT*, pp. 265-269, 1997.

[17] P.Sca lart and J. Vieira-Filho, "Speech enhancement based on a priori signal to noise estimation," *Proc. ICASSP*, pp. 629-632, 1996.

[18] R. McAullay and M. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, pp. 137-145, 1980.

[19] R. McAullay and M. Malpass, "Speech enhancement using a soft-decision maximum liklihood noise suppression filter," Tech. note 1979-31, M.I.T. Lincoln Lab., Lexington, MA, June 1979.

[20] B. L. Sim, Y. C. Tong, J. S. Chang and C. T. Tan, "A Parametric formulation of the generalized spectral subtraction method," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 328-337, 1998.

[21] E. Nemer, R. Goubran and S. Mahmoud, "Speech enhancement using fourth-order cumulants and optimum filters in the subband domain," *Speech Communication*, vol. 36, pp. 219-246, 2002.

[22] C. Avendano, H. Hermansky, M. Vis and A. Bayya, "Adative speech enhancement using frequency specific SNR estimates," *Proceedings of III IEEE Workshop on Interactive voice Technology for Telecommunications Applications*, Basking Ridge, New Jersey, pp. 65-68, 1996

[23] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 251-266, 1995.

[24] J. Huang and Y. Zhao, "An energy-constrained signal subspace method for speech enhancement and recognition in white and colored noises," Speech Communication, vol. 26, 1998, pp. 165-181.

[25] Y. Ephraim , "A Bayesian estimation approach for speech enhancement using hidden Markov model," *IEEE Trans. Sinal Processing*, vol. 40, pp. 725-735, 1992.

[26] H. Sameti, H. Sheikhzadeh, L. Deng and R. L. Brennan, "HMM-based strategies for enhancement of speech signals embedded in nonstationary noise," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 5, pp. 445-455, 1998.

[27] B. Yegnanarayana, C. Avendano, H. Hermansky and P. S. Murthy, "Speech enhancement using linear prediction residual," *Speech Communication*, vol. 28, pp. 25-42, 1999.

[28] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-26, pp. 197-210, 1978.

[29] R. McAullay and M. Malpass, "Estimation of noise corrupted speech DFT-spectrum using the pitch period," *IEEE Trans. Speech Audio Processing*, vol. 2, pp. 1-8, 1994.

[30] I. Y. Soon, S. N. Koh and C. K. Yeo, "Noisy speech enhancement using discrete cosine transform," *Speech Communication*, vol. 24, pp. 249-257, 1998.

[31] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 613-627, May 1995.

[32] J. W. Seok and K. S. Bae, "Speech enhancement with reduction of noise components in the wavelet domain," in *Proc. of ICASSP*, pp. II-1323-1326, 1997.

[33] D. Mahmoudi, "A microphone array for speech enhancement using multiresolution wavelet transform," in *Proc. Eurospeech'97*, Rhodes, Greece, pp. 339-342, 1997.

[34] T. Gulzow, A. Engelsberg and U. Heute, "Comparison of a discrete wavelet transformation and nonuniform polyphase filter bank applied to spectral-subtraction speech enhancement," *Signal Process.*, vol. 64, pp. 5-19, 1998.

[35] J. Sika and V. Davidek, "Multi-channel noise reduction using wavelet filter bank," in *Proc. Eurospeech'97*, Rhodes, Greece, 1997.

[36] D. Mahmoudi and A. Drygajlo, "Combined Weiner and coherence filtering in wavlet domain for microphone array speech enhancement," in *ICASSP*, Seattle, WA, pp. 358-388, 1998.

[37] M. Bahoura and J. Rouat, "Wavelet speech enhancement based on the teager energy operator," *IEEE Signal Processing Letters*, vol. 8, no. 1, pp. 10-12, January 2001.

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

[39] D. L. Wang and J. S. Lim, "The unimportance of phase in speech enhancement," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 30, pp. 679-681, 1982.

[40] M. R. Weiss, E. Aschkenasy and T. W. Parsons, "Processing of speech signals to attenuate interference," in *IEEE Symp. Speech Recognition*, (Pittsburgh, PA), pp. 292-293, 1974.

[41] N. Virag, "Single channel speech enhnacement system based on masking properties of the human auditory system," *IEEE Trans. Speech Audio Processing*, vol. 7, pp. 126-137, 1999.

[42] H. Gustaffson, S. E. Nordholm and I. Claesson, "Spectral subtraction using reduced delay convolution and adaptive averaging," *IEEE Trans. Speech, Audio Process.*, vol. 9, pp. 799-807, 2001.

[43] I. Y. Soon and S. N. Koh, "Low distortion speech enhancement," *IEE Proc.-Vis. Image Signal Process.*, vol. 147, pp.247-253, 2000.

[44] I. Y. Soon, S. N. Koh and C. K. Yeo, "Improved noise suppression filter using self-adaptive estimator of probability of speech absence," *Signal Processing 75 (1999), 151-159.*

[45] N. S. Kim and J. H. Chang, "Spectral enhancement based on global soft decision," *IEEE Signal Process. Letters*, vol. 7 No. 5, May 2000.

[46] J. Jensen and J. H. L. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," *IEEE Trans. on Speech and Audio Processing*, vol. 9, no. 7, pp. 731-740, Oct. 2001.

Appendix A

Derivation of $\alpha_{n,k}$ in the FFT Domain

In the FFT domain

$$\begin{aligned} Y &= X + D \\ &= X_R + jX_I + D_R + jD_I \end{aligned} \tag{A.1}$$

where X_R and X_I are real and imaginary parts of $X_{n,k}$, respectively, and D_R and D_I are real and imaginary parts of $D_{n,k}$, respectively. $|Y|^2$ can be written as

$$\begin{aligned} |Y|^2 &= (X_R^2 + D_R^2) + (X_I^2 + D_I^2) \\ &\quad +2X_R D_R + 2X_I D_I \\ &= |X|^2 + |D|^2 + 2X_R D_R + 2X_I D_I \end{aligned} \tag{A.2}$$

Thus $E\{|Y|^4\}$ is obtained as

$$\begin{aligned} E\{|Y|^4\} &= E\{|X|^4 + |D|^4 + 2|X|^2|D|^2 + 4(X_R^2 D_R^2 + X_I^2 D_I^2 \\ &\quad +2X_R X_I D_R D_I) + 4(|X|^2 + |D|^2)(X_R D_R + X_I D_I)\} \\ &= E\{|X|^4\} + E\{|D|^4\} + 2E\{|X|^2\}E\{|D|^2\} \\ &\quad +4\left(\frac{1}{2}E\{|X|^2\}\right)\left(\frac{1}{2}E\{|D|^2\}\right) \\ &\quad +4\left(\frac{1}{2}E\{|X|^2\}\right)\left(\frac{1}{2}E\{|D|^2\}\right) \end{aligned} \tag{A.3}$$

Using the fact that $X_{n,k}$ and $D_{n,k}$ are zero-mean and uncorrelated complex Gaussian random variables, we can assume $E\{X_R X_I D_R D_I\} = 0$, $E\{|X|^2(X_R D_R + X_I D_I)\} = 0$, $E\{|D|^2(X_R D_R + X_I D_I)\} = 0$, $E\{|X_R|^2\} = E\{|X_I|^2\} = \frac{1}{2}E\{|X|^2\}$, $E\{|D_R|^2\} = E\{|D_I|^2\} = \frac{1}{2}E\{|D|^2\}$, thus simplified form of $E\{|Y|^4\}$ is obtained

as follows

$$E\{|Y|^4\} = E\{|X|^4\} + E\{|D|^4\} + 4E\{|X|^2\}E\{|D|^2\} \quad (A.4)$$

Again introducing notational subscript (n,k)

$$E\{|Y_{n,k}|^4\} = E\{|X_{n,k}|^4\} + E\{|D_{n,k}|^4\} + 4E\{|X_{n,k}|^2\}E\{|D_{n,k}|^2\} \quad (A.5)$$

Since $X_{n,k}$ is a complex Gaussian random process, the probability density function of $|X_{n,k}|$ follows the Rayleigh distribution, i.e.,

$$p(|X_{n,k}|) = \frac{2|X_{n,k}|}{E\{|X_{n,k}|^2\}} \exp\left(-\frac{|X_{n,k}|^2}{E\{|X_{n,k}|^2\}}\right) \quad (A.6)$$

Fourth moment of $|X_{n,k}|$, i.e. $E\{|X_{n,k}|^4\}$, is then given by

$$\begin{aligned} E\{|X_{n,k}|^4\} &= \int_0^\infty |X_{n,k}|^4 p(|X_{n,k}|) d|X_{n,k}| \\ &= \frac{2}{E\{|X_{n,k}|^2\}} \int_0^\infty |X_{n,k}|^5 \exp\left(-\frac{|X_{n,k}|^2}{E\{|X_{n,k}|^2\}}\right) d|X_{n,k}| \end{aligned} \quad (A.7)$$

Using Eq. (2.33) $E\{|X_{n,k}|^4\}$ is obtained as

$$E\{|X_{n,k}|^4\} = 2E\{|X_{n,k}|^2\}^2 \quad (A.8)$$

Similarly,

$$E\{|D_{n,k}|^4\} = 2E\{|D_{n,k}|^2\}^2 \quad (A.9)$$

Substituting values from Eqs. (A.8) and (A.9) into Eq. (A.5)

$$\begin{aligned} E\{|Y_{n,k}|^4\} &= 2E\{|X_{n,k}|^2\}^2 + 2E\{|D_{n,k}|^2\}^2 \\ &\quad +4E\{|X_{n,k}|^2\}E\{|D_{n,k}|^2\} \end{aligned} \quad (A.10)$$

In the FFT domain, the local *a posteriori* SNR and *a priori* SNR are defined as follows:

$$\mathrm{SNR}_{post}(n,k) = \gamma_{n,k} = \frac{|Y_{n,k}|^2}{\sigma_d^2(n,k)} \quad (A.11)$$

$$\mathrm{SNR}_{prior}(n,k) = \xi_{n,k} = \frac{E\{|X_{n,k}|^2\}}{\sigma_d^2(n,k)} \quad (A.12)$$

where $\sigma_d^2(n,k) = E\{|D_{n,k}|^2\}$. Using Eqs. (A.10) and (A.11), we get

$$E\{(\gamma_{n,k}-1)^2\} = \frac{E\{|Y_{n,k}|^4\}}{E\{|D_{n,k}|^2\}^2} - 2E\{\gamma_{n,k}\} + 1$$

$$
\begin{aligned}
&= \frac{2E\{|X_{n,k}|^2\} + 2E\{|D_{n,k}|^2\} + 4E\{|X_{n,k}|^2\}E\{|D_{n,k}|^2\}}{E\{|D_{n,k}|^2\}^2} \\
&\quad -2E\{\gamma_{n,k}\} + 1 \\
&= 2\frac{E\{|X_{n,k}|^2\}^2}{E\{|D_{n,k}|^2\}^2} + 2\frac{E\{|D_{n,k}|^2\}^2}{E\{|D_{n,k}|^2\}^2} + 4\frac{E\{|X_{n,k}|^2\}E\{|D_{n,k}|^2\}}{E\{|D_{n,k}|^2\}^2} \\
&\quad -2E\{\gamma_{n,k}\} + 1 \\
&= 2\xi_{n,k}^2 + 2 + 4\xi_{n,k} - 2E\{\gamma_{n,k}\} + 1 \qquad \text{(A.13)}
\end{aligned}
$$

As $E\{(\gamma_{n,k} - 1)\} = \xi_{n,k}$, it follows that $E\{\gamma_{n,k}\} = 1 + \xi_{n,k}$ [[15], Eq. (49)]. Thus Eq. (A.13) becomes

$$
\begin{aligned}
E\{(\gamma_{n,k} - 1)^2\} &= 2\xi_{n,k}^2 + 4\xi_{n,k} + 2 - 2(1 + \xi_{n,k}) + 1 \\
&= 2\xi_{n,k}^2 + 2\xi_{n,k} + 1 \qquad \text{(A.14)}
\end{aligned}
$$

Substituting $E\{(\gamma_{n,k} - 1)\} = \xi_{n,k}$ and $E\{(\gamma_{n,k} - 1)^2\} = 2\xi_{n,k}^2 + 2\xi_{n,k} + 1$ into Eq. (3.12), we obtain

$$
\begin{aligned}
J_\alpha &= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + (1 - \alpha_{n,k})^2(2\xi_{n,k}^2 + 2\xi_{n,k} + 1) \\
&\quad +2\alpha_{n,k}(1 - \alpha_{n,k})\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}\xi_{n,k}\tilde{\xi}_{n-1,k} \\
&\quad -2(1 - \alpha_{n,k})\xi_{n,k}\xi_{n,k} + \xi_{n,k}^2 \\
&= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + 2\xi_{n,k}^2(1 - \alpha_{n,k})^2 + 2\xi_{n,k}(1 - \alpha_{n,k})^2 + (1 - \alpha_{n,k})^2 \\
&\quad +2\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}\xi_{n,k}\tilde{\xi}_{n-1,k} \\
&\quad -2(1 - \alpha_{n,k})\xi_{n,k}\xi_{n,k} + \xi_{n,k}^2 \\
&= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + 2\xi_{n,k}^2(1 - \alpha_{n,k})^2 + 2\xi_{n,k}(1 - \alpha_{n,k})^2 + (1 - \alpha_{n,k})^2 \\
&\quad +2\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} - 2\alpha_{n,k}\tilde{\xi}_{n-1,k}\xi_{n,k} \\
&\quad -2(1 - \alpha_{n,k})\xi_{n,k}^2 + \xi_{n,k}^2 \\
&= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \xi_{n,k}^2\left[2(1 - \alpha_{n,k})^2 - 2(1 - \alpha_{n,k}) + 1\right] \\
&\quad +\xi_{n,k}\left[2(1 - \alpha_{n,k})^2 + 2\alpha_{n,k}\tilde{\xi}_{n-1,k} - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k} - 2\alpha_{n,k}\tilde{\xi}_{n-1,k}\right] \\
&\quad +(1 - \alpha_{n,k})^2 \\
&= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \xi_{n,k}^2(2 - 4\alpha_{n,k} + 2\alpha_{n,k}^2 - 2 + 2\alpha_{n,k} + 1) \\
&\quad +\xi_{n,k}\left[2(1 - \alpha_{n,k})^2 - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\right] + (1 - \alpha_{n,k})^2 \\
&= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \xi_{n,k}^2(1 - 2\alpha_{n,k} + 2\alpha_{n,k}^2) + \xi_{n,k}\left[2(1 - \alpha_{n,k})^2 - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k})\right] \\
&\quad +(1 - \alpha_{n,k})^2 \\
&= \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \alpha_{n,k}^2\xi_{n,k}^2 + \xi_{n,k}^2(1 - 2\alpha_{n,k} + \alpha_{n,k}^2)
\end{aligned}
$$

$$
\begin{aligned}
& -2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} + 2\xi_{n,k}(1-\alpha_{n,k})^2 + (1-\alpha_{n,k})^2 \\
= \; & \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 + \alpha_{n,k}^2\xi_{n,k}^2 + \xi_{n,k}^2(1-\alpha_{n,k})^2 - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} \\
& +2\xi_{n,k}(1-\alpha_{n,k})^2 + (1-\alpha_{n,k})^2 \\
= \; & \alpha_{n,k}^2\tilde{\xi}_{n-1,k}^2 - 2\alpha_{n,k}^2\tilde{\xi}_{n-1,k}\xi_{n,k} + \alpha_{n,k}^2\xi_{n,k}^2 + \xi_{n,k}^2(1-\alpha_{n,k})^2 \\
& +2\xi_{n,k}(1-\alpha_{n,k})^2 + (1-\alpha_{n,k})^2 \\
= \; & \alpha_{n,k}^2(\tilde{\xi}_{n-1,k}^2 - 2\tilde{\xi}_{n-1,k}\xi_{n,k} + \xi_{n,k}^2) + (1-\alpha_{n,k})^2(\xi_{n,k}^2 + 2\xi_{n,k} + 1) \\
= \; & \alpha_{n,k}^2(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + (1-\alpha_{n,k})^2(\xi_{n,k}+1)^2 \qquad \text{(A.15)}
\end{aligned}
$$

Differentiating J_α with respect to $\alpha_{n,k}$ gives

$$
\begin{aligned}
\frac{\partial J_\alpha}{\partial \alpha_{n,k}} &= 2\alpha_{n,k}(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 2(1-\alpha_{n,k})(-1)(\xi_{n,k}+1)^2 \\
&= 2\alpha_{n,k}(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 2\alpha_{n,k}(\xi_{n,k}+1)^2 - 2(\xi_{n,k}+1)^2 \\
&= \alpha_{n,k}\left(2(\tilde{\xi}_{n-1,k} - \xi_{n,k})^2 + 2(\xi_{n,k}+1)^2\right) - 2(\xi_{n,k}+1)^2 \qquad \text{(A.16)}
\end{aligned}
$$

Now equating $\partial J_\alpha / \partial \alpha_{n,k}$ to zero yields an optimum expression for $\alpha_{n,k}$ as

$$
\alpha_{n,k}^{opt} = \frac{1}{1 + \left(\dfrac{\xi_{n,k} - \tilde{\xi}_{n-1,k}}{\xi_{n,k}+1}\right)^2} \qquad \text{(A.17)}
$$