SPEECH ENHANCEMENT USING A LAPLACIAN-BASED MMSE ESTIMATOR OF THE MAGNITUDE SPECTRUM

APPROVED BY SUPERVISORY COMMITTEE:

Dr. Philipos C. Loizou, Chair

Dr. Mohammad Saquib

Dr. Issa M S Panahi

Dr. Hlaing Minn

© Copyright 2005
Bin Chen
All Rights Reserved

To My Wife Yan.

SPEECH ENHANCEMENT USING A LAPLACIAN-BASED MMSE ESTIMATOR OF THE MAGNITUDE SPECTRUM

by

BIN CHEN, B.S.E.E., M.S.E.E.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN ELECTRICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT DALLAS

December 2005

ACKNOWLEDGEMENTS

My first thank you is to Dr. Philipos C. Loizou, my advisor, for his many suggestions and constant support during my Ph.D studies. I am very grateful for those research opportunities he provided me in the past four years and I believe this research experience will greatly benefit my career in the future. I learned from Dr. Loizou that persistent effort and faith are undoubtedly the most important assets. This enables me to cast a new light on traditional approaches of speech enhancement and improve them by proposing our methods. I believe our work in the past several years advanced the state of art of the speech enhancement problem. My greatest gratitude to him!

I am also very thankful to my committee members Dr. Saquib, Dr. Minn and Dr. Panahi for their precious time and their valuable comments and suggestions in my thesis proposal defense. They dramatically improved my final dissertation. I took one course taught by each of the three professors who are great teachers and very good at clarifying confusing concepts. I learned much in their classes, especially in DSP II and in discussions with Dr. Panahi. I would also like to mention that Dr. Minn's guidance in the early phases of the dissertation work along with constant consultations with Dr. Saquib were indispensable to the completion of my Ph.D work.

Head of mathematics department, Dr. Hooshyar, was very helpful in my research

work and supplied me with his expertise about the numerical methods such as efficient implementations of numerical integration, numerical stability analysis and convergence conditions.

I had the pleasure of working with my lab mates Yi, Felipe, Kalyan and Sundar. My friends Feng and Ruiming also gave me a lot of help in my PhD studies. They are wonderful people.

I should also acknowledge that my PhD studies were supported in part by National Institutes of Health. Of course, I am so grateful to my parents who guided me from a skip-class kid to a self-disciplined young man, and encouraged me to pursue more advanced education in the United States. Without them I would never have come so far in pursuing my career and dream.

Finally, I would say to my wife Yan: "It is so fortunate for me to have such a wonderful wife as you and for your constant understanding, support and love. Without your sacrifice this project would never have been accomplished and this dissertation would never have come into existence." Now that I am finishing my Ph.D studies soon, you have just begun steps to your degree and I wish you success in your qualifying exams.

Richardson, Texas *Bin Chen*

November 18, 2005

SPEECH ENHANCEMENT USING A LAPLACIAN-BASED MMSE ESTIMATOR OF THE MAGNITUDE SPECTRUM

Publication No. _____

Bin Chen, Ph.D.

The University of Texas at Dallas, 2005

Supervising Professor: Dr. Philipos C. Loizou

A number of speech enhancement algorithms based on MMSE spectrum estimators have been proposed over the years. Although some of these algorithms were developed based on Laplacian and Gamma distributions, no optimal *spectral magnitude* estimators were derived. This dissertation focuses on optimal estimators of the magnitude spectrum for speech enhancement. We present an analytical solution for estimating in the MMSE sense the magnitude spectrum when the clean speech DFT coefficients are modeled by a Laplacian distribution and the noise DFT coefficients are modeled by a Gaussian distribution. Furthermore, we derive the MMSE estimator under speech presence uncertainty and a Laplacian statistical model. Results indicated that the Laplacian-based MMSE estimator yielded less residual noise in the enhanced speech than the traditional

Gaussian-based MMSE estimator. Overall, the present study demonstrates that the assumed distribution of the DFT coefficients can have a significant effect on the quality of the enhanced speech.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

In general, there exists a need for digital voice communications, human-machine interfaces, and automatic speech recognition systems to perform reliably in noisy environments. For example, in hands-free operation of cellular phones in vehicles, the speech signal to be transmitted may be contaminated by reverberation and background noise. In many cases, these systems work well in nearly noise-free conditions, but their performance deteriorates rapidly in noisy conditions. Therefore, development of preprocessing algorithms for speech enhancement is always of interest. The goal of speech enhancement varies according to specific applications, such as to boost the overall speech quality, to increase intelligibility, and to improve the performance of voice communication devices [2]. Recent research also shows promising applications to cochlear implants [3].

Most single-channel speech enhancement algorithms rely on frequency domain weighting [4] [1] [5] [6] [7], commonly consisting of a noise spectral density estimator [8] [9] [10] and a spectral amplitude estimator [1] [11] [12]. However, the non-stationary nature of speech signals and limitation on algorithmic delay and complexity require that the noisy signal must be processed in short frames. Based on that requirement, some speech estimators apply a statistical estimation rule based on a statistical model of the Discrete Fourier Transform (DFT) coefficients. The well-known examples are Minimum

Mean Square Error (MMSE) algorithms [1] [13] [14] and Wiener filtering [15] [2] [16], which estimate the complex speech DFT coefficients based on Minimum Mean Square Error (MMSE).

1.1 Problem statement

A number of frequency domain speech enhancement algorithms were proposed in [4] [17] [18] [8] [19] [20]. This dissertation focuses on Short-Time Spectral Amplitude (STSA) estimators using the MMSE criterion, which have received a lot of attention in the past two decades [1] [13] [21] [22]. In the previous work, Ephraim and Malah [1] optimally estimated, under an assumed Gaussian statistical model, the STSA and complex exponential of the phase of the speech signal. They used this approach to optimally estimate the short-time spectral amplitude (STSA), rather than the real and imaginary components of the Short-Time Fourier Transform (STFT) separately, since the STSA of a speech signal rather than its waveform is of major importance in speech perception. It has been shown that the STSA and the complex exponential cannot be estimated simultaneously in an optimal way. Therefore, an optimal MMSE STSA estimator was employed and combined with the phase obtained from the noisy spectrum.

The assumed Gaussian statistical model provides a good approximation for the noise DFT coefficients [23]. For speech signals, however, whose typical DFT frame sizes used in mobile communications are short (10ms -40ms) that assumption is not well fulfilled. It is valid only if the DFT frame size is much longer than the span of

correlation of the signal under consideration [24]. It has been shown that the real and imaginary part of the speech coefficients are better modeled with Laplacian and Gamma densities [25] [12] [26]. The observation led researchers to derive a similar optimal MMSE STSA estimator but based on more accurate models, Laplacian and/or Gamma. The derivation of such an estimator is complicated leading some people to seek alternative techniques to compute the MMSE STSA estimator.

1.2 Proposed Work

In this work, an analytical solution of the optimal MMSE estimator obtained from a direct derivation of STSA is presented based on Laplacian modeling and a mild approximation. In order to further improve the performance of the estimator, we also derive a Speech Presence Uncertainty (SPU) estimator and incorporate it into the amplitude estimator. In detail, we target the following problems:

1. We derive the joint probability density function (pdf) of the magnitude and phase of the DFT coefficients, and use that to derive the MMSE estimator of the speech magnitude spectrum based on a Laplacian model for the speech DFT coefficients and a Gaussian model for the noise DFT coefficients. After making some assumptions (and approximations) about the relationship between the magnitude and phase pdfs (probability density functions), we derive a Laplacian-based MMSE estimator in closed form.

2. In order to view how the assumption about the relationship between the magnitude and phase pdfs affects the estimator, we also derive its expression (no

solution obtained) without any assumptions, and compare the performance of this estimator to the performance of the approximate estimator.

3. Speech Presence Uncertainty (SPU) estimator is another contribution to the proposed estimator. Unlike Voice Activity Detection (VAD), it addresses the speech/silence detection problem in terms of probability, and is derived using the Bayes' rule. After the magnitude is estimated by the method proposed above, the SPU estimator refines the estimate of the magnitudes by scaling them by the SPU probability.

This thesis is organized as follows. Chapter 2 gives a comprehensive review for the state-of-the-art MMSE methods in speech enhancement. The proposed MMSE methods with/without assumptions are presented and compared, and related numerical stability are discussed in Chapter 3. In Chapter 4, the Speech Presence Uncertainty (SPU) estimator is described. Chapter 5 presents the results of the proposed work and comparison with other methods. Numerical stability and implementation issues are further discussed and compared in this chapter. Chapter 6 gives a summary of our contributions.

CHAPTER 2
LITERATURE REVIEW

In this chapter, we present the literature on MMSE estimators used for speech enhancement.

2.1 Overview of the MMSE approach

In frequency-domain speech enhancement, the magnitude spectrum of the speech signal is estimated and combined with the short-time phase of the degraded speech to produce the enhanced signal. The spectral subtraction algorithms [27] [28] [3] and Wiener filtering [15] [2] [16] are well-known examples. In the spectral subtraction algorithms, the STSA of noisy and noise signals are estimated as the square root of maximum likelihood estimation of each signal spectral variance [16], and then subtracted from the magnitude spectrum of the noisy signal. In the Wiener filtering algorithm, the estimator is obtained by finding the optimal MMSE estimates of the complex Fourier transform coefficients. Both spectrum subtraction and Wiener filtering algorithms are derived under Gaussian assumption for each spectral component.

Spectrum subtraction and Wiener filtering are not optimal *spectral amplitude* estimators under the assumed Gaussian model. As stated in Chapter 1, the spectral amplitude estimators are more advantageous than spectral estimators from a perceptual

point of view. Ephraim and Malah [1] [13] formulated an optimal spectral amplitude estimator, which, specifically, estimates the modulus (magnitude) of each complex Fourier coefficient of the speech signal in a given analysis frame from the noisy speech in that frame.

In order to derive the MMSE STSA estimator, the *a priori* probability distribution of the speech and noise Fourier expansion coefficients should be assumed since these are unknown in reality. Ephraim-Malah [1] derived their spectral amplitude estimator based on a Gaussian model. They argued that this assumption utilizes asymptotic statistical properties of the Fourier coefficients. Specifically, according to [29] [30], they assumed that the Fourier expansion coefficients of each process can be modeled as statistically independent Gaussian random variables, real- and imaginary parts of each component is independent to each other, and the mean of each coefficient is assumed to be zero and the variance time-varying. The assumption is motivated by the central limit theorem [31]. Central limit theorem/asymptotic statistical properties hold more strongly when the analysis frame size is long, which somewhat conflicts with the "short-time" requirement.

Porter and Boll [25], Brehm and Stammler [32], and Martin [12] recognized that the DFT coefficients of clean speech derived from speech frames having a length of about the span of correlation within the signal are not normally distributed, and might be better modeled by a Gamma or a Laplacian distribution when the DFT frame length is in the range of 10-40 ms [25] [12] [26], which is a typical frame size in speech applications. He provided histograms of DFT coefficients using his own experimental

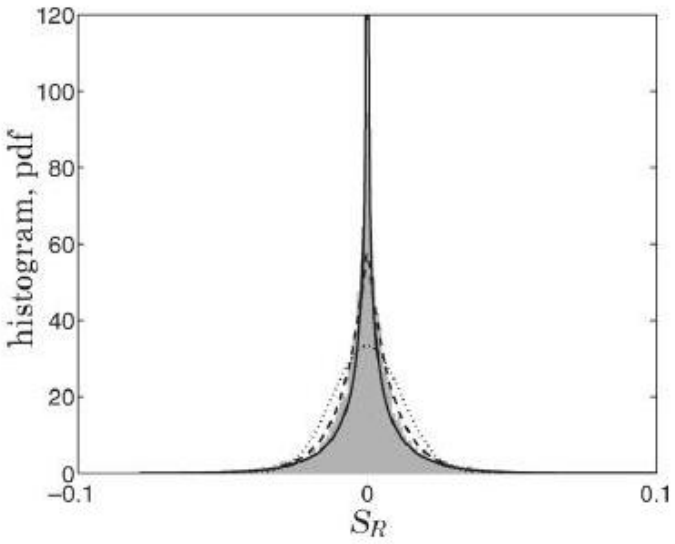

Figure 2.1. (Dotted) Gaussian, (dashed) Laplacian, and (solid) Gamma densities fitted to a histogram (shaded) of the real part of clean speech DFT coefficients ($L = 256$ $f_s = 8000$Hz)

data to support the argument. Figures 2.1 and 2.2 from [33] show the histograms.

Figures 2.1 and 2.2 plot the histograms of the real part of the DFT coefficients (frame size $L = 256$, sampling frequency = 8000 Hz) of clean speech averaged over three male and three female speakers and six minutes of speech. At a global SNR of more than 40 dB, only a very low, almost stationary, level of quantization and ambient recording noise is present in these signals. Since speech signals are highly non-stationary, it must be ensured that the spectral coefficients represented in the histogram are obtained under quasi-stationary conditions. Martin [33] tried to guarantee the stationarity as follows. The *a priori* SNR can be used to select DFT coefficients with a consistent quasi-stationary SNR, or, in the case of stationary background noise, with a fixed power. That is, power is used as a criterion for stationarity. So for the depicted histograms,

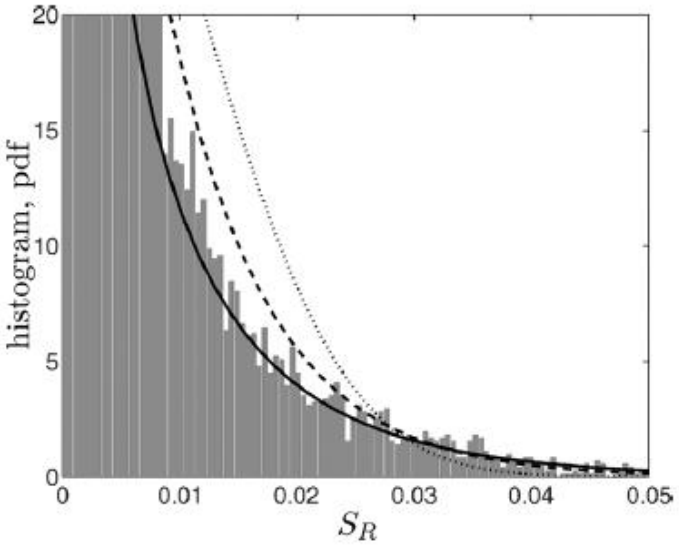

Figure 2.2. (Dotted) Gaussian, (dashed) Laplacian, and (solid) Gamma densities fitted to a histogram (shaded) of the real part of clean speech DFT coefficients ($L = 256$ $f_s = 8000$Hz)

only those DFT coefficients are selected which have an estimated *a priori* SNR larger than 28 dB and smaller than 30 dB. Martin [33] also extended his investigation on model assumptions to car noise. His results showed that this kind of noise does not necessarily follow a Gaussian distribution as shown in Figures 2.3 and 2.4.

Figures 2.3 and 2.4 from [33] plot the histogram and the corresponding Gaussian and Laplacian model densities for highly-correlated car noise (Diesel engine) recorded at a constant speed of 90 km/h and evaluated with a DFT length of $L = 256$ and sampling frequency of 8000Hz. It was found that the Laplacian density provided a reasonable fit to the experimental data. Breithaupt and Martin [26] further verified that the Gamma pdf produced a better fit to experimental data than the Gaussian pdf using the Kullback divergence measure [34].

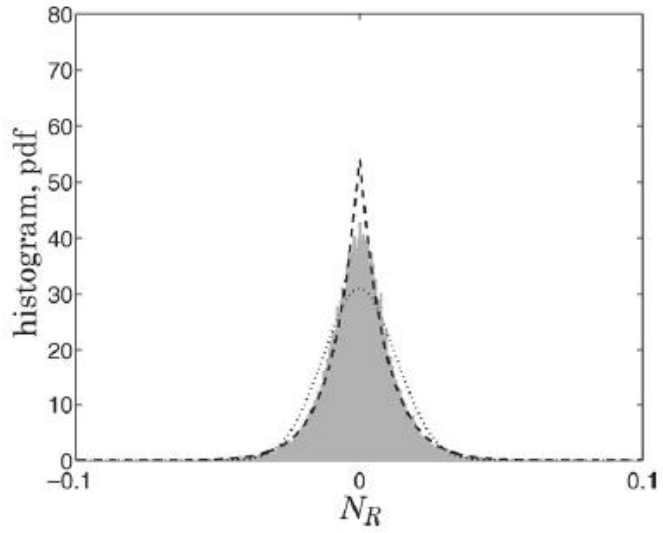

Figure 2.3. (Dotted) Gaussian and (dashed) Laplacian densities fitted to a histogram (shaded) of the real part of car noise DFT coefficients ($L = 256$ $f_s = 8000$Hz)

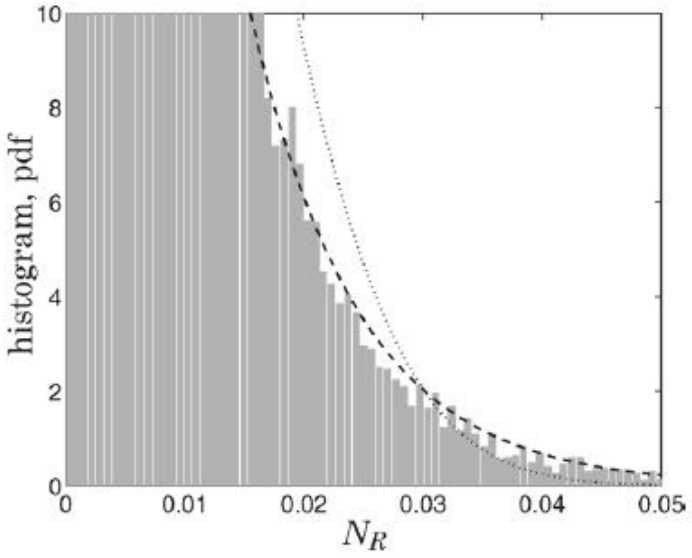

Figure 2.4. (Dotted) Gaussian and (dashed) Laplacian densities fitted to a histogram (shaded) of the real part of car noise DFT coefficients ($L = 256$ $f_x = 8000$Hz)

Breithaupt and Martin [26] proposed MMSE Squared Spectral Magnitude (SSM) Estimators based on Gamma and Laplacian models. In their approach, they aimed to compute the squared magnitude of the clean speech spectral coefficients by estimating the squared real and imaginary parts of the DFT respectively. In their method, the clean speech signal and noise were modeled by a combination of Gaussian, Gamma and Laplacian distributions, which will be detailed later.

It is known that signal enhanced by the MMSE estimator of the power spectral density [26] suffers from musical noise. Martin [35] proposed a new estimator, in which the real and imaginary parts of the clean signal were estimated in the MMSE sense conditional on the real and imaginary parts of the observed noisy signal. This estimator, however, is not the optimal *spectral amplitude* estimator, which, we have pointed out in Chapter 1, is more preferable perceptually.

Lotter and Vary [36] addressed the issue of finding the optimal *spectral amplitude* estimator based on Gamma or Laplacian models. They found a closed form of the spectral amplitude estimator, but they had to employ a parametric function to approximate the pdf of the spectral amplitude since the analytic solution of the pdf was unknown to them. The parametric function had a simpler form and facilitated their derivation but the authors did not specify how they obtained the parametric function, and the parameters needed to be determined empirically. The following sections will give more details about the methods mentioned above.

2.2 The Gaussian based MMSE STSA Estimator

Let $y(n) = x(n) + d(n)$ be the sampled noisy speech signal consisting of the clean signal $x(n)$ and the noise signal $d(n)$. Taking the short-time Fourier transform of $y(n)$, we get:

$$Y(\omega_k) = X(\omega_k) + D(\omega_k) \tag{2.1}$$

for $\omega_k = 2\pi k/N$ where $k = 0, 1, 2, ...N-1$, and N is the frame length. The above equation can also be expressed in polar form as

$$Y_k e^{j\theta_y(k)} = X_k e^{j\theta_x(k)} + D_k e^{j\theta_d(k)} \tag{2.2}$$

For convenience, we denote the modulus $|X(\omega_k)|$ as X_k or x_k in the rest of the dissertation. Since the spectral components are assumed to be statistically independent, the MMSE amplitude estimator $\hat{X}_k$ can be derived from $Y(\omega_k)$ only. That is,

$$\begin{aligned} \hat{X}_k &= E\{X_k | Y(\omega_0), Y(\omega_1), ...\} \\ &= E\{X_k | Y(\omega_k)\} \\ &= \frac{\int_0^\infty \int_0^{2\pi} x_k p(Y(\omega_k)|x_k, \theta_k) p(x_k, \theta_k) d\theta_k dx_k}{\int_0^\infty \int_0^{2\pi} p(Y(\omega_k)|x_k, \theta_k) p(x_k, \theta_k) d\theta_k dx_k} \end{aligned} \tag{2.3}$$

where $\theta_k = \theta_x(k)$.

Under the assumed Gaussian model, $p(Y(\omega_k)|x_k, \theta_k)$ and $p(x_k, \theta_k)$ are given by [1]

$$p(Y(\omega_k)|x_k, \theta_k) = \frac{1}{\pi\lambda_d(k)} \exp\left\{-\frac{1}{\lambda_d(k)} \left|Y_k - X_k e^{j\theta_x(k)}\right|^2\right\} \tag{2.4}$$

$$p(x_k, \theta_k) = \frac{x_k}{\pi\lambda_x(k)} \exp\left\{-\frac{X_k^2}{\lambda_x(k)}\right\} \tag{2.5}$$

where $\lambda_x(k) \triangleq E\{|X_k|^2\}$, and $\lambda_d(k) \triangleq E\{|D_k|^2\}$, are the variances of the kth spectral component of the speech and the noise respectively. Substituting Eq. (2.4) and Eq. (2.5) into Eq. (2.3) gives [1]

$$\begin{aligned} \hat{X}_k &= \Gamma(1.5)\frac{\sqrt{\upsilon_k}}{\gamma_k} M(-0.5;1;-\upsilon_k) R_k \\ &= \Gamma(1.5)\frac{\sqrt{\upsilon_k}}{\gamma_k} \exp\left(-\frac{\upsilon_k}{2}\right)\left[(1+\upsilon_k) I_0\left(\frac{\upsilon_k}{2}\right) + \upsilon_k I_1\left(\frac{\upsilon_k}{2}\right)\right] Y_k \end{aligned} \tag{2.6}$$

where $\Gamma(\cdot)$ denotes the gamma function, with $\Gamma(1.5) = \sqrt{\pi}/2$; $M(a;c;x)$ is the confluent hypergeometric function defined as

$$\begin{aligned} M(a;c;x) &= \sum_{r=0}^{\infty} \frac{(a)_r}{(c)_r}\frac{x^r}{r!} \\ &= 1 + \frac{a}{c}\frac{x}{1!} + \frac{a(a+1)}{c(c+1)}\frac{x^2}{2!} + ... \end{aligned}$$

where $(a)_r \triangleq a(a+1)\cdots(a+r-1)$, and $(a)_0 \triangleq 1$. $I_0(\cdot)$ and $I_1(\cdot)$ denote the modified Bessel functions of zero and first order, respectively, υ_k is defined by

$$\upsilon_k \triangleq \frac{\xi_k}{1+\xi_k}\gamma_k \tag{2.7a}$$

where ξ_k and γ_k are defined by

$$\xi_k \triangleq \frac{\lambda_x(k)}{\lambda_d(k)} \tag{2.7b}$$

$$\gamma_k \triangleq \frac{Y_k^2}{\lambda_d(k)} \tag{2.7c}$$

ξ_k and γ_k are interpreted as the *a priori* and *a posteriori* signal-to-noise ratio (SNR), respectively. Ephraim-Malah [1] also showed that at high SNR, i.e. $\xi_k \gg 1$ and $\upsilon_k \gg 1$,

$$M(-0.5;1;-\upsilon_k) \simeq \frac{\sqrt{\upsilon_k}}{\Gamma(1.5)} \qquad \upsilon_k \gg 1 \tag{2.8}$$

and the estimator can be simplified as

$$\hat{X}_k \triangleq \frac{\xi_k}{1+\xi_k} Y_k \tag{2.9}$$

which is the Wiener estimator. Note that the confluent hypergeometric function $M(a;c;x)$ can be efficiently calculated using [37].

2.3 The Laplacian/Gamma based MMSE Estimator of the Power Spectrum

Breithaupt and Martin [26] proposed their MMSE SSM in terms of the following combinations of statistical models:

1. Gaussian Noise and Gaussian Speech Model
2. Gaussian Noise and Gamma Speech Model
3. Laplacian Noise and Gamma Speech Model

The MMSE estimation of $|X_k|^2$ is obtained by the conditional expectation $E\{|X_k|^2|Y(\omega_k)\}$. Since it is assumed that the real and imaginary parts of a DFT coefficient are independent and identically distributed(i.i.d), the estimation can be split into an estimator for the real part $X_{kR} = Re\{X(\omega_k)\}$ and an estimator for the imaginary part X_{kI} =Imag$\{X(\omega_k)\}$. The estimator then becomes:

$$E\{|X_k|^2|Y(\omega_k)\} = E\{|X_{kR}|^2|Y_{kR}\} + E\{|X_{kI}|^2|Y_{kI}\} \tag{2.10}$$

The three distributions considered were:

1. Gaussian pdf

$$p_x(x) = \frac{1}{\sqrt{\pi\lambda_x}} \exp\left(-\frac{x^2}{\lambda_x}\right) \tag{2.11}$$

2. Laplacian pdf

$$p_x(x) = \frac{1}{2\sqrt{\lambda_x}} \exp\left(-\frac{|x|}{\sqrt{\lambda_x}}\right) \tag{2.12}$$

3. Gamma pdf

$$p_x(x) = \frac{\sqrt[4]{3}}{2\sqrt[4]{2\lambda_x}\sqrt{\pi}} |x|^{-1/2} \exp\left(-\frac{\sqrt{3}}{\sqrt{2}}\frac{|x|}{\sqrt{\lambda_x}}\right) \tag{2.13}$$

.

Different combinations of the above pdfs were considered in [26] yielding different estimators.

2.3.1 Gaussian Noise and Gaussian Speech Model

The MMSE estimator of the power spectral density $E\{|X_k|^2|Y(\omega_k)\}$ has been computed in [38] and is given by:

$$E\{|X_k|^2|Y(\omega_k)\} = \frac{\xi_k}{1+\xi_k}\lambda_d(k) + \left(\frac{\xi_k}{1+\xi_k}Y_k\right)^2 \tag{2.14}$$

where ξ_k is the *a priori* SNR defined earlier.

2.3.2 Gaussian Noise and Gamma Speech Model

The following estimator was found when modeling the pdf of noise with a Gaussian distribution and the pdf of speech by a Gamma distribution:

$$E\{|X_{kR}|^2|Y_{kR}\} = \frac{3}{16}\lambda_d(k)\frac{\Psi\left(\frac{5}{4},\frac{1}{2};G_+^2\right) + \Psi\left(\frac{5}{4},\frac{1}{2};G_-^2\right)}{\Psi\left(\frac{1}{4},\frac{1}{2};G_+^2\right) + \Psi\left(\frac{1}{4},\frac{1}{2};G_-^2\right)} \tag{2.15}$$

where

$$G_{\pm} = \frac{\sqrt{3}}{2\sqrt{2}\sqrt{\xi_k}} \pm \frac{|Y_{kR}|}{\sqrt{\lambda_d(k)}} \tag{2.16}$$

If we define

$$\begin{aligned} Den &= \frac{4\sqrt{\pi}}{\Gamma(\frac{1}{4})} G_{-} \Phi\left(\frac{3}{4}, \frac{3}{2}; -G_{-}^2\right) \\ &\quad - \exp(-G_{-}^2)\left[\Psi\left(\frac{1}{4}, \frac{1}{2}; G_{+}^2\right) + \Psi\left(\frac{5}{4}, \frac{1}{2}; G_{-}^2\right)\right] \end{aligned} \tag{2.17}$$

we have

$$\begin{aligned} E\{|X_{kR}|^2|Y_{kR}\} &= \frac{3}{16}\lambda_d(k) \cdot \frac{1}{Den} \cdot \left\{\frac{16\sqrt{\pi}}{\Gamma(\frac{1}{4})} G_{-} \Phi\left(-\frac{1}{4}, \frac{3}{2}; -G_{-}^2\right)\right. \\ &\quad \left. - \exp(-G_{-}^2)\left[\Psi\left(\frac{5}{4}, \frac{1}{2}; G_{+}^2\right) + \Psi\left(\frac{5}{4}, \frac{1}{2}; G_{-}^2\right)\right]\right\} \end{aligned} \tag{2.18}$$

where $\Phi(\alpha; \gamma; x) = F_1(\alpha; \gamma; x)$ denotes the confluent hypergeometric function and $\Psi(\alpha; \gamma; x)$ is defined in [39].

2.3.3 Laplacian Noise and Gamma Speech Model

Assuming a Laplacian pdf for the noise DFT coefficients, and a Gamma pdf for the speech DFT coefficients, we get the MMSE estimator of the squared real part of the clean speech DFT coefficient as follows:

$$\begin{aligned} E\{|X_{kR}|^2|Y_{kR}\} &= \frac{\lambda_d(k)}{4G_{+}^2} \cdot \frac{1}{Den} \cdot \left\{3\sqrt{\pi}\right. \\ &\quad +4\exp\left(-G_{-}\frac{|Y_R|}{\sqrt{\lambda_d(k)}}\right)\Psi\left(-\frac{3}{2}, -\frac{3}{2}; G_{+}\frac{|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right) \\ &\quad \left. +\frac{8}{5}\left(G_{+}\frac{|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right)^{5/2}\Phi\left(\frac{8}{5}, \frac{7}{2}; -G_{-}\frac{|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right)\right\} \end{aligned} \tag{2.19}$$

where

$$G_{\pm} = \frac{\sqrt{3}}{\sqrt{2}\sqrt{\xi_k}} \pm 2 \tag{2.20}$$

and

$$\begin{aligned} Den &= \sqrt{\pi} + \exp(-G_-^2 \frac{|Y_R|}{\sqrt{\lambda_d(k)}}) \Psi\left(\frac{1}{2}, \frac{1}{2}; G_+ \frac{|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right) \\ &\quad +2\left(G_+ \frac{|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right)^{1/2} \Phi\left(\frac{1}{2}, \frac{3}{2}; -G_- \frac{|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right) \end{aligned}$$

With the signal modeled by a Gamma pdf, the authors [26] obtained consistently better results than those modeled by Gaussian pdf. However, musical noise was introduced with the Gaussian/Gamma estimator. That is probably due to inaccuracy of power spectrum estimation. A detailed discussion about the relation between *a priori SNR* and musical noise in MMSE based algorithms can be found in [1] [40]. A potential improvement with the multitaper method was presented in [41] [19] [42] but the authors did not compare their work to it.

2.4 The Laplacian/Gamma based Complex Spectrum MMSE Estimator

Following the approach in [26], Martin [35] also derived complex spectrum MMSE estimators assuming the following pdfs for the speech and noise DFT coefficients:

1. Gaussian Noise and Gaussian Speech Model

2. Gaussian Noise and Laplacian Speech Model

3. Laplacian Noise and Speech Model

For each of the above three combinations of pdfs, Martin [35] found an analytical solution for estimating complex DFT coefficients in the MMSE sense. In his derivation, he also assumed that the real and imaginary parts of a DFT coefficient were independent and identically distributed(i.i.d). The estimation can be split into an estimator for the real part $X_{kR} = Re\{X(\omega_k)\}$ and an estimator for the imaginary part X_{kI} =Imag$\{X(\omega_k)\}$. The estimator then becomes

$$E\{X(\omega_k)|Y(\omega_k)\} = E\{X_{kR}|Y_{kR}\} + jE\{X_{kI}|Y_{kI}\} \tag{2.21}$$

The MMSE complex spectrum estimators based on the above three combinations of modeling are briefly introduced in the following sections.

2.4.1 Gaussian Noise and Gaussian Speech Model

When both noise and speech DFT coefficients are assumed to have complex Gaussian pdfs, the optimal MMSE estimator is the Wiener filter [43]:

$$\hat{X}_k = E\{X(\omega_k)|Y(\omega_k)\} = \frac{\lambda_x(k)}{\lambda_x(k) + \lambda_d(k)} Y_k = \frac{\xi_k}{1 + \xi_k} Y_k \tag{2.22}$$

where $\lambda_x(k) \triangleq E\{X_k^2\}$, $\lambda_d(k) \triangleq E\{D_k^2\}$, and ξ_k denotes the a priori SNR at frequency bin k.

2.4.2 Gaussian Noise and Laplacian Speech Model

Assuming now that the prior of speech DFT coefficients is Laplacian distributed and the noise is Gaussian distributed, we get:

$$E\{X_{kR}|Y_{kR}\} = \frac{\sqrt{\lambda_d(k)}\left[L_{R_+} \exp\left(L_{R_+}^2\right) erfc(L_{R_+}) - L_{R_-} \exp\left(L_{R_-}^2\right) erfc(L_{R_-})\right]}{\exp\left(L_{R_+}^2\right) erfc(L_{R_+}) + \exp\left(L_{R_-}^2\right) erfc(L_{R_-})} \tag{2.23}$$

$$E\{X_{kI}|Y_{kI}\} = \frac{\sqrt{\lambda_d(k)}\left[L_{I_+}\exp\left(L_{I_+}^2\right)erfc(L_{I_+}) - L_{I_-}\exp\left(L_{I_-}^2\right)erfc(L_{I_-})\right]}{\exp\left(L_{I_+}^2\right)erfc(L_{I_+}) + \exp\left(L_{I_-}^2\right)erfc(L_{I_-})} \quad (2.24)$$

where $erfc(\cdot)$ is the error function [44] and

$$L_{kR_+} = \frac{1}{\sqrt{\xi_k}} + \frac{Y_{kR}}{\sqrt{\lambda_d(k)}}$$

$$L_{kR_-} = \frac{1}{\sqrt{\xi_k}} - \frac{Y_{kR}}{\sqrt{\lambda_d(k)}}$$

$$L_{kI_+} = \frac{1}{\sqrt{\xi_k}} + \frac{Y_{kI}}{\sqrt{\lambda_d(k)}}$$

$$L_{kI_-} = \frac{1}{\sqrt{\xi_k}} - \frac{Y_{kI}}{\sqrt{\lambda_d(k)}}$$

2.4.3 Laplacian Noise and Speech models

When both the real and the imaginary components of the noise and the speech coefficients are modeled by a Laplacian pdf, the optimal estimator of the real part is given by

$$\begin{aligned} E\{X_{kR}|Y_{kR}\} = & \frac{sign(Y_{kR})}{\exp\left(\frac{-2|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right)\sqrt{\lambda_d(k)} - \exp\left(\frac{-2|Y_{kR}|}{\sqrt{\lambda_x(k)}}\right)\sqrt{\lambda_x(k)}} \\ & \cdot\frac{\lambda_x(k)\lambda_d(k)}{\lambda_d(k)-\lambda_x(k)}\left(\exp\left(\frac{-2|Y_{kR}|}{\sqrt{\lambda_d(k)}}\right) - \exp\left(\frac{-2|Y_{kR}|}{\sqrt{\lambda_x(k)}}\right)\right) \\ & -|Y_{kR}|\exp\left(\frac{-2|Y_{kR}|}{\sqrt{\lambda_x(k)}}\right)\sqrt{\lambda_x(k)} \end{aligned}$$

Although separate treatment of the DFT components simplifies the derivation to some degree, the estimators appear to be complicated. Experimental results confirmed that the above MMSE estimators performed significantly better than the squared spectrum magnitude estimators [26] in terms of segmental SNR.

2.5 MAP STSA Estimator using a parametric approximation function

As mentioned earlier, a spectral amplitude estimator is more preferable than a complex *spectrum* estimator. Lotter and Vary [36] proposed a MMSE Short Time Spectral Amplitude (STSA) estimator using a parametric function to approximate the pdf of the spectral amplitude. The pdf of the spectral amplitude was approximated by the following parametric function:

$$p(X_k) = \frac{\mu^{\alpha+1}}{\Gamma(\alpha+1)} \frac{X_k^{\alpha}}{\lambda_x(k)^{\frac{\alpha+1}{2}}} \exp\left\{-\mu \frac{X_k}{\sqrt{\lambda_x(k)}}\right\}$$

where α and μ are parameters. The parametric function was shown to match the histogram of $p(X_k)$ (the true spectral amplitude pdf) best when $\alpha = 1$ and $\mu = 2.5$.

Using the above parametric approximation function, Lotter and Vary [36] sought a Maximum *a Posterior* (MAP) estimator, i.e.,

$$\hat{X}_k = \arg\max_{X_k} \{p(X_k|Y_k)\} = \arg\max_{X_k} \left\{\frac{p(Y_k|X_k)p(X_k)}{p(Y_k)}\right\} \tag{2.25}$$

Since $p(Y_k)$ is independent of X_k, we can maximize:

$$p(Y_k|X_k)p(X_k) = {X_k}^{\alpha-\frac{1}{2}} \exp\left\{-\frac{X_k^2}{\delta_N^2} - X_k\left(\frac{\mu}{\sqrt{\lambda_x(k)}} - \frac{2Y_k}{\lambda_d(k)}\right)\right\} \tag{2.26}$$

After taking the log of $p(Y_k|X_k)p(X_k)$, and differentiating it with respect to X_k, we get:

$$\frac{d\log\left[p(Y_k|X_k)p(X_k)\right]}{dX_k} = \left(\alpha - \frac{1}{2}\right)\frac{1}{X_k} - \frac{2X_k}{\lambda_d(k)} - \frac{\mu}{\sqrt{\lambda_x(k)}} + \frac{2Y_k}{\lambda_d(k)} = 0$$

Solving for X_k, we get the following MAP estimator:

$$\hat{X}_k = Y_k\left(u + \sqrt{u^2 + \frac{\alpha - \frac{1}{2}}{2\gamma_k}}\right) \tag{2.27}$$

where $u = \frac{1}{2} - \frac{\mu}{4\sqrt{\gamma\xi}}$

The MAP estimator was evaluated with 3 kinds of noise: white, ventilator, and cafeteria noise. Better noise reduction was achieved with the MAP estimator compared to the MMSE estimator [1].

2.6 Summary

The MMSE estimation of speech spectrum have received considerable research attention [1] [13] [21] [22]. Although significant progress has been made in the above mentioned methods, each of them has its own weaknesses and limitations either on the underlying assumptions or derivation of the estimators. The fact that Laplacian and Gamma are better than the Gaussian pdf for modeling the speech DFT coefficients has been recognized and empirically verified by many researchers. Therefore, many have focused on Laplacian/Gamma based estimators. None of them, however, has succeeded in finding the preferred optimal *spectral amplitude* estimator in the MMSE sense.

CHAPTER 3

LAPLACIAN-BASED MMSE STSA ESTIMATOR

3.1 Introduction

The basic idea of Laplacian based MMSE STSA estimator is to find the optimal estimate of the modulus of the speech signal DFT components in the MMSE sense, based on the assumption that the real and imaginary parts of these components are modeled by a Laplacian distribution. The noise signal DFT components are assumed to be Gaussian distributed. Note that the assumption about noise DFT component is valid regardless of whether the noise is white or colored. One of the challenges in deriving such an estimator is to compute the pdf of spectral amplitude. It is especially complicated when the real and imaginary parts of the DFT coefficients are modeled by a Laplacian distribution. The Laplacian assumption makes the derivation in question even more difficult in that independence between amplitude and phase no longer holds with the Laplacian distribution.

In this chapter, we derive two Laplacian-based MMSE STSA estimators. The first estimator makes no assumptions or approximations. The second estimator makes a mild assumption about the independence of spectral amplitude and phase. This assumption enables us to derive the MMSE estimator in closed form.

3.2 True Laplacian Based Estimator

In this section, we derive the Laplacian based estimator making no assumptions about the independence of amplitude and phase.

3.2.1 Principles

According to the last chapter, the MMSE magnitude estimator is given as:

$$\begin{aligned} \hat{X}_k &= E\{X_k|Y(\omega_k)\}, \quad k = 0, 1, 2, ...N-1 \\ &= \frac{\int_0^\infty \int_0^{2\pi} x_k p(Y(\omega_k)|x_k, \theta_k) p(x_k, \theta_k) d\theta_k dx_k}{\int_0^\infty \int_0^{2\pi} p(Y(\omega_k)|x_k, \theta_k) p(x_k, \theta_k) d\theta_k dx_k} \end{aligned} \tag{3.1}$$

where $p(Y(\omega_k)|x_k, \theta_k)$ is the conditional pdf of noisy spectrum and is given by:

$$p(Y(\omega_k)|x_k, \theta_k) = \frac{1}{\pi\lambda_d(k)} \exp\left\{-\frac{1}{\lambda_d(k)}|Y(\omega_k) - X(\omega_k)|^2\right\} \tag{3.2}$$

and $p(x_k, \theta_k)$ is the joint pdf of the magnitude spectrum and phases. Following the procedure in [45], it is easy to show that $p(x_k, \theta_k)$ is given by:

$$p(x_k, \theta_k) = \frac{x_k}{2\sqrt{\lambda_x(k)}} \exp\left[-\frac{x_k}{\sqrt{\lambda_x(k)}}\left(|\cos\theta_k| + |\sin\theta_k|\right)\right] \tag{3.3}$$

Substituting Eq. (3.2) and Eq. (3.3) into Eq. (3.1), we get:

$$\begin{aligned} \hat{X}_k &= \frac{\int_0^\infty \int_0^{2\pi} x_k^2 \exp\left[-\frac{Y_k^2 - 2x_k Re\{e^{-j\theta_x} Y(\omega_k)\} + X_k^2}{\lambda_d(k)}\right] \exp\left[\frac{-x_k}{\sqrt{\lambda_x(k)}}\left(|\cos\theta_k| + |\sin\theta_k|\right)\right] d\theta_x dx_k}{\int_0^\infty \int_0^{2\pi} x_k \exp\left[-\frac{Y_k^2 - 2x_k Re\{e^{-j\theta_x} Y(\omega_k)\} + X_k^2}{\lambda_d(k)}\right] \exp\left[\frac{-x_k}{\sqrt{\lambda_x(k)}}\left(|\cos\theta_k| + |\sin\theta_k|\right)\right] d\theta_x dx_k} \\ &= \frac{\int_0^\infty x_k^2 \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) \int_0^{2\pi} \exp\left(\frac{2x_k \cos\theta_k Y(\omega_k)}{\lambda_d(k)} - \frac{x_k|\cos\theta_k|}{\sqrt{\lambda_x(k)}} - \frac{x_k|\sin\theta_k|}{\sqrt{\lambda_x(k)}}\right) d\theta_x dx_k}{\int_0^\infty x_k \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) \int_0^{2\pi} \exp\left(\frac{2x_k \cos\theta_k Y(\omega_k)}{\lambda_d(k)} - \frac{x_k|\cos\theta_k|}{\sqrt{\lambda_x(k)}} - \frac{x_k|\sin\theta_k|}{\sqrt{\lambda_x(k)}}\right) d\theta_x dx_k} \\ &= \frac{\int_0^\infty x_k^2 \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) \int_0^{2\pi} \exp\left[\left(\frac{2x_k Y(\omega_k)}{\lambda_d(k)} \pm \frac{x_k}{\sqrt{\lambda_x(k)}}\right)\cos\theta_k \pm \frac{x_k}{\sqrt{\lambda_x(k)}}\sin\theta_k\right] d\theta_x dx_k}{\int_0^\infty x_k \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) \int_0^{2\pi} \exp\left[\left(\frac{2x_k Y(\omega_k)}{\lambda_d(k)} \pm \frac{x_k}{\sqrt{\lambda_x(k)}}\right)\cos\theta_k \pm \frac{x_k}{\sqrt{\lambda_x(k)}}\sin\theta_k\right] d\theta_x dx_k} \end{aligned} \tag{3.4}$$

Note that we have 4 cases ($\pm$X$\pm$) to evaluate since we have $|\cos\theta_k|$ and $|\sin\theta_k|$ in $[0 \;\; 2\pi]$. For $\int_0^{2\pi} \exp\left[\left(\frac{2x_kY(\omega_k)}{\lambda_d(k)} \pm \frac{x_k}{\sqrt{\lambda_x(k)}}\right)\cos\theta_k \pm \frac{x_k}{\sqrt{\lambda_x(k)}}\sin\theta_k\right] d\theta_x$, we transform $A\cos\theta_k \pm B\sin\theta_k$ into $\sqrt{A^2+B^2}\cos(\theta_k \mp \arccos\frac{A}{\sqrt{A^2+B^2}})$ first and then integrate it. There is, however, the difficulty that $A = \frac{2x_kY(\omega_k)}{\lambda_d(k)} \pm \frac{x_k}{\sqrt{\lambda_x(k)}}$ and $B = \frac{x_k}{\sqrt{\lambda_x(k)}}$ are determined by experimental data since the $\lambda_d(k)$, noise variance, and $\lambda_x(k)$, signal variance, are involved. The derivation can not go on with a variable angle $\arccos\frac{A}{\sqrt{A^2+B^2}}$ determined empirically because we can not determine the integral limits.

We were not able to find the closed form solution of the estimator in Eq. (3.4), which we will be referring to as the LapMMSE estimator. Hence, we took an alternative approach, and made some assumptions about the independence of magnitude and phase.

3.2.2 Derivation of pdf of spectral amplitude with complex Laplacian distributed DFT coefficients

It is known that if X_R and X_I are Gaussian distributed, then $X = \sqrt{X_R^2 + X_I^2}$ is Rayleigh distributed. The distribution of X is not known when X_R and X_I are *Laplacian* distributed. We, therefore, focused on deriving the pdf of the spectral amplitude with complex Laplacian distributed DFT coefficients. The pdf of the spectral amplitude is given by (see Appendix A):

$$p_X(x) = \frac{2x}{\lambda}\int_0^{\pi/4} \exp(-\frac{\sqrt{2}}{\sqrt{\lambda}}x\cos\theta)d\theta \tag{3.5}$$

where λ denotes the variance. Solving the integral in Eq. (A.13) (see Appendix A), we get the resulting pdf:

$$p_X(x) = \frac{\pi x}{2\lambda} I_0\left(-\frac{\sqrt{2}}{\sqrt{\lambda}}x\right) + \frac{4x}{\lambda}\sum_{n=1}^{\infty}\frac{1}{n} I_n\left(-\frac{\sqrt{2}}{\sqrt{\lambda}}x\right)\sin\frac{\pi n}{4}, \quad x \geqslant 0 \tag{3.6}$$

where $I_n(\cdot)$ denotes the modified Bessel function of nth order [39]. In the following derivations, $\lambda_x(k)$ denotes the variance of signal x at frequency bin k. Similarly, X_k denotes the spectral magnitude of signal x at frequency bin k.

Simulations were run to verify that the derived pdf of the spectral amplitude X_k is correct. Figure 3.1 shows these simulations, where the "+" line represents the plot of the function $p_X(x) = \frac{2x}{\sigma^2}\int_0^{\pi/4}\exp(-\frac{\sqrt{2}}{\delta}x\cos\theta)d\theta$ by numerical computation, while the solid line represents the plot from its solution $\frac{\pi x}{2\lambda}I_0\left(-\frac{\sqrt{2}}{\sqrt{\lambda}}x\right) + \frac{4x}{\lambda}\sum_{n=1}^{\infty}\frac{1}{n}I_n\left(-\frac{\sqrt{2}}{\sqrt{\lambda}}x\right)\sin\frac{\pi n}{4}$. We can see that the two plots are almost the same. The small difference is due to the truncation of the infinite summation in Eq. (3.6) to the first 40 terms.

Let us consider the joint pdf of the magnitude spectrum and phases, $p(x_k, \theta_k)$. The marginal pdf $p(x_k)$ is clearly the pdf of the spectral amplitude X_k and is given by Eq. (3.6). The marginal pdf $p(\theta_k)$ is given by:

$$\begin{aligned} p(\theta_k) &= \frac{x_k}{2\sigma}\int_0^{\infty}\exp\left[-\frac{x_k}{\sigma}(|\cos\theta_k| + |\sin\theta_k|)\right]dx_k \\ &= -\frac{x_k}{2|\cos\theta_k| + 2|\sin\theta_k|}\exp\left[-\frac{x_k}{\sigma}(|\cos\theta_k| + |\sin\theta_k|)\right]/_0^{\infty} \\ &\quad -\frac{\sigma}{2 + 4|\cos\theta_k\sin\theta_k|}\exp\left[-\frac{x_k}{\sigma}(|\cos\theta_k| + |\sin\theta_k|)\right]/_0^{\infty} \\ &= \frac{\sigma}{2 + 4|\cos\theta_k\sin\theta_k|} \end{aligned} \tag{3.7}$$

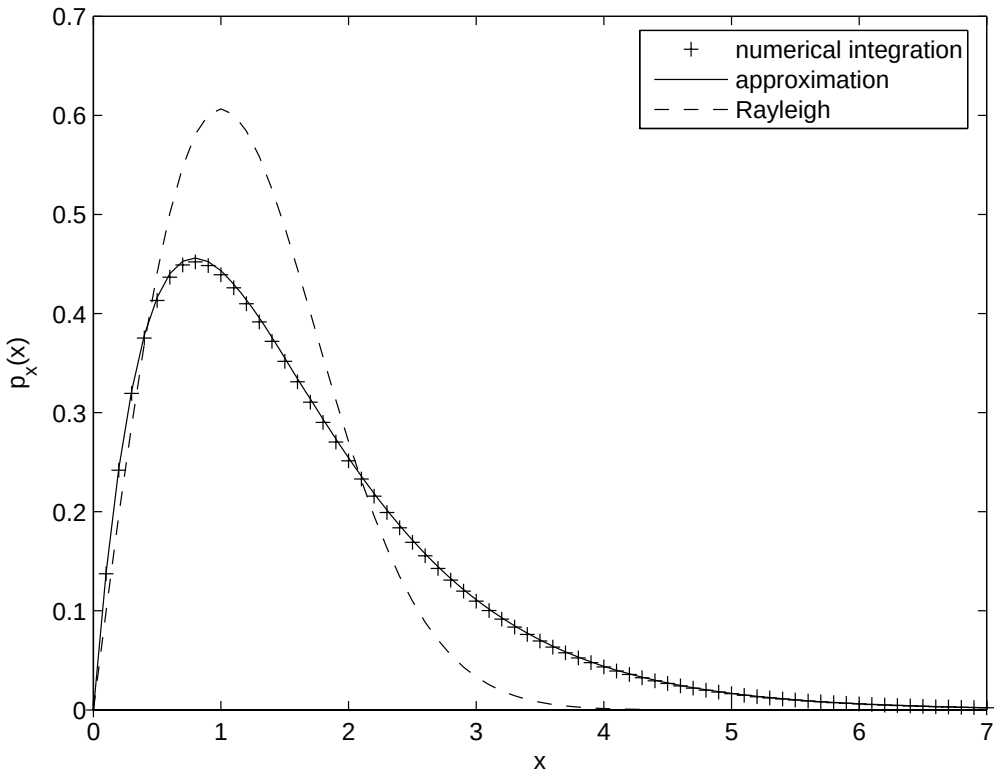

Figure 3.1. The pdf of the magnitude DFT coefficients assuming the real and imaginary parts are modeled by a Laplacian distribution ($\sigma^2 = 1$). The plot indicated by '+' shows the pdf computed by numerical integration of Eq. (A.5). The plot indicated by the solid line shows the pdf approximated by truncating the infinite summation in Eq. (3.6) with the first 40 terms. The Rayleigh distribution (dashed line), used in the Gaussian-based MMSE estimator [1], is superimposed for comparative purposes.

The joint pdf $p(x_k, \theta_k)$ is given by:

$$p(x_k, \theta_k) = \frac{x_k}{2\sqrt{\lambda_x(k)}} \exp\left[-\frac{x_k}{\sqrt{\lambda_x(k)}}\left(|\cos\theta_k| + |\sin\theta_k|\right)\right] \tag{3.8}$$

It is clear that $p(x_k, \theta_k) \neq p(x_k) \cdot p(\theta_k)$, hence the amplitude is not independent of the phase. The fact that X_k and θ_k are not independent complicates the derivation of the MMSE estimator (see Eq. (3.4)). Hence, we examined whether X_k and θ_k were "nearly" independent.

3.3 Approximate Laplacian MMSE estimator

It is known that complex zero mean Gaussian random variables have magnitudes and phases which are statistically independent [45]. Furthermore, the phases have a uniform distribution. This is not the case, however, with the complex Laplacian distributions proposed in this dissertation for modeling the speech DFT coefficients. Further analysis of the joint pdf of the magnitudes and phases, $p(x_k, \theta_k)$, however, revealed that the pdfs of the magnitudes and phases are nearly statistically independent, at least for a certain range of magnitude values. To show that, we derived the marginal pdfs of the magnitudes and phases and examined whether $p(x_k, \theta_k) \approx p(x_k)p(\theta_k)$.

The marginal pdf of the amplitudes and the phases have been given in the last section. Figures 3.2 and 3.3 show plots of the joint density $p(x_k, \theta_k)$ as well as plots of the product of the magnitude and phase pdfs. Figure 3.2 shows the joint density $p(x_k, \theta_k)$, Figure 3.3 shows $p(x_k)p(\theta_k)$ and Figure 3.4 shows the absolute difference between the densities displayed in Figures 3.2 and 3.3. As can be seen, the difference

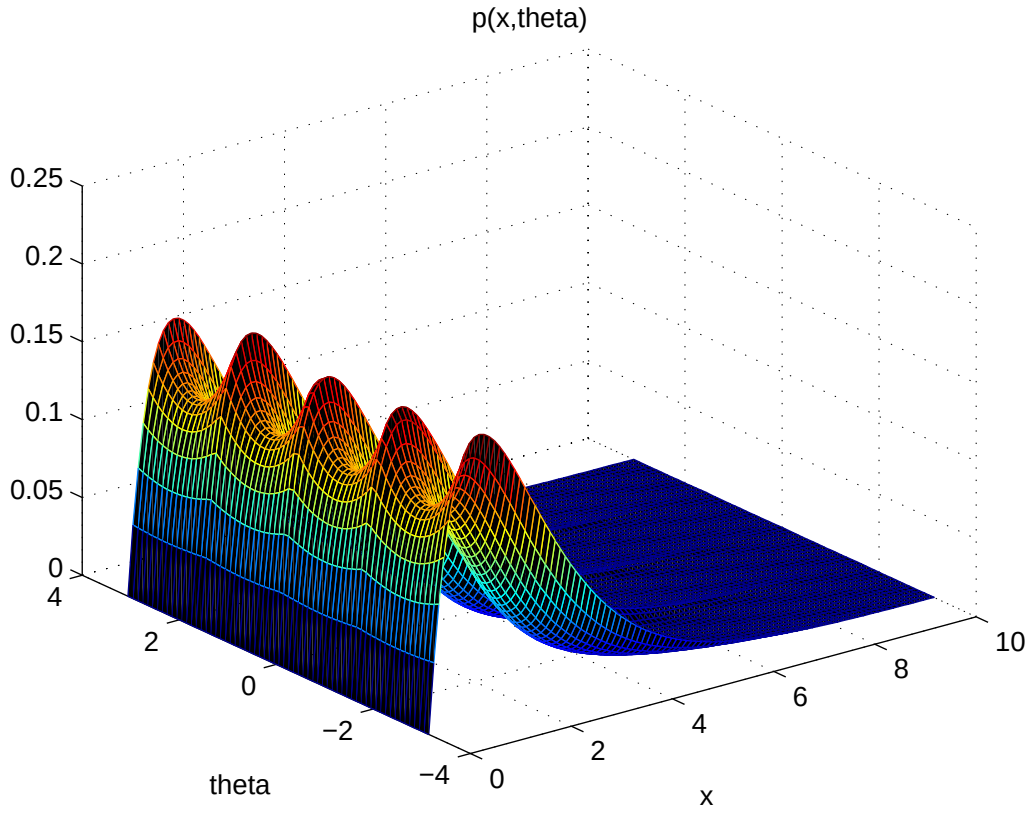

Figure 3.2. Plot of the joint density $p(x_k, \theta)$ of a zero mean complex Laplacian random variable ($\sigma^2 = 1$).

between the two densities is large near $x_k \approx 0$, but is near zero for $x_k > 2$. The plot in Figure 3.4 demonstrates that the magnitudes and phases are nearly independent, at least for a specific range of magnitude values ($x_k > 2$, $\theta_k \in [-\pi, \pi]$). We can therefore make the approximation that the joint density $p(x_k, \theta_k) \approx p(x_k)p(\theta_k)$.

We further analyzed the phase pdf, $p(\theta_k)$, to determine the shape of the distribution and examine whether it is similar to a uniform distribution. Figure 3.5 shows the plot of $p(\theta_k)$ along with a uniform distribution. The density $p(\theta_k)$ is clearly not uniform, but it has characteristics similar to a uniform distribution, i.e., it is relatively

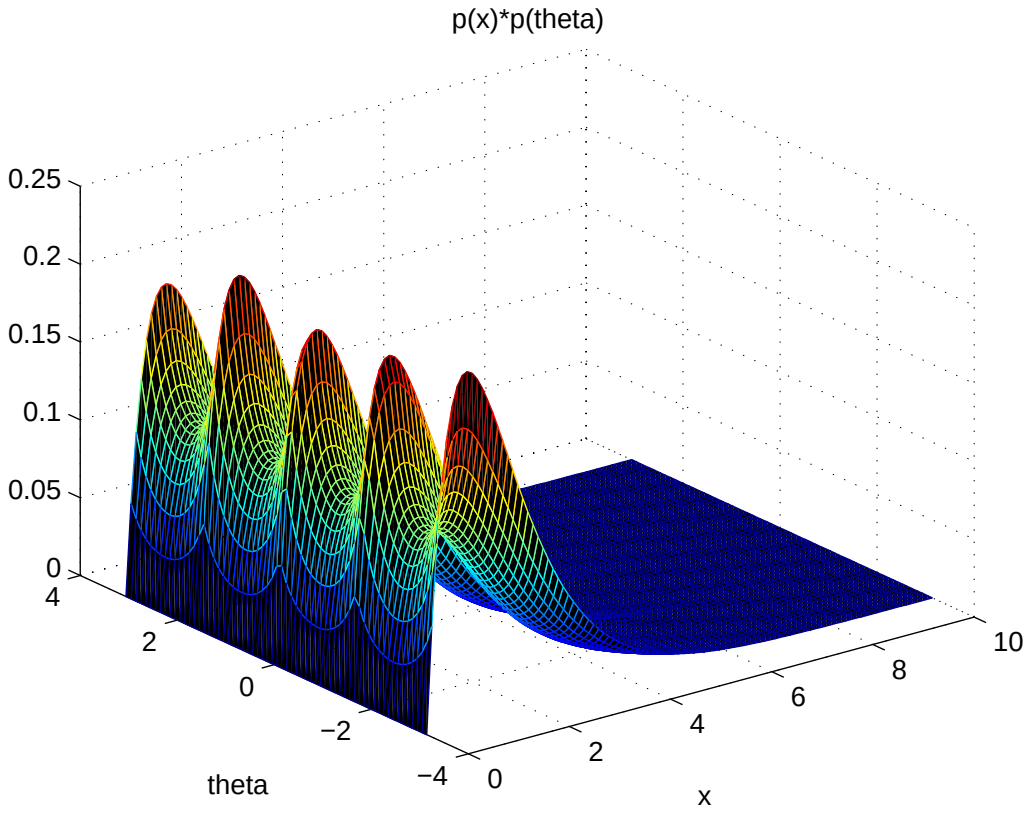

Figure 3.3. Plot of $p(x_k)p(\theta)$, where $p(x_k)$ is given by Eq. (3.6) and $p(\theta)$ is given by Eq. (3.7) $[\lambda_x(k) = 1]$.

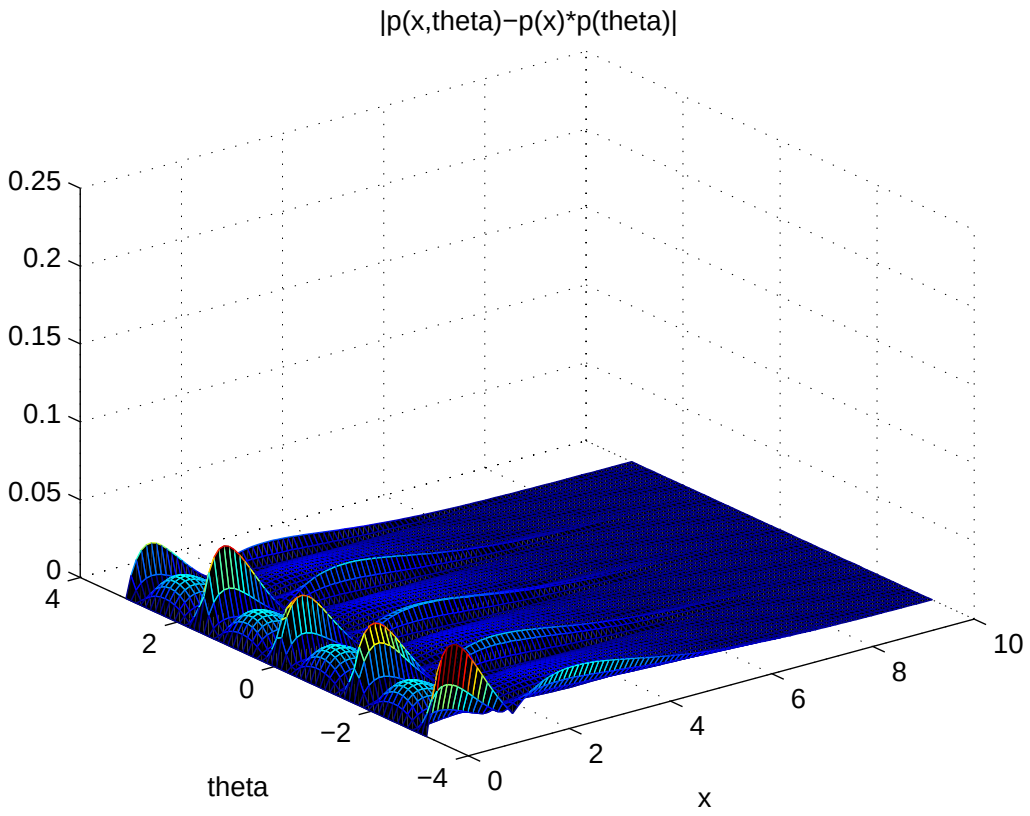

Figure 3.4. Plot of the absolute difference between the densities shown in Fig. 3.2 and 3.3.

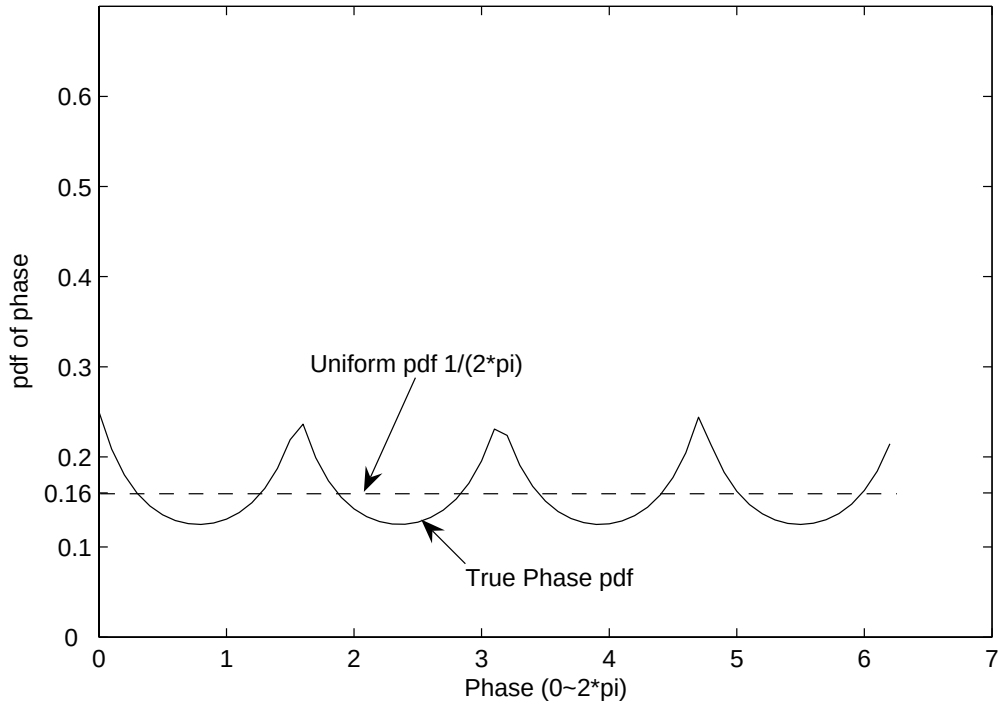

Figure 3.5. Plot of $p(\theta)$ (solid lines) superimposed to a uniform distribution (dashed lines).

flat for $\theta_k \in [-\pi, \pi]$. We therefore approximated $p(\theta_k)$ with a uniform distribution, i.e., $p(\theta_k) \approx 1/(2\pi)$ for $\theta_k \in [-\pi, \pi]$.

Based on the above two assumptions (statistical independence between x_k and θ_k, and a uniform distribution for the phases), we approximated the joint density in Eq. (3.1) with $p(x_k, \theta_k) \approx \frac{1}{2\pi} p(x_k)$, where $p(x_k)$ is the density of the spectral magnitudes. After substituting Eq. (3.2) and Eq. (3.6) into Eq. (3.1), we obtain the MMSE estimator in closed form as follows (see Appendix B for detailed derivation):

$$\hat{X}_k = \frac{A_k + B_k}{C_k + D_k} \tag{3.9a}$$

where

$$A_k = \frac{\left(\frac{Y_k^2}{\gamma_k}\right)^{\frac{3}{2}}}{2} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{3}{2})}{m!\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m \cdot F\left(-m, -m; 1; 2\xi_k^2 Y_k^2\right) \tag{3.9b}$$

$$\begin{aligned} B_k \;=\; & \frac{8}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin\frac{\pi n}{4} \frac{\left(\frac{2\gamma_k}{Y_k}\right)^n \left(\frac{\gamma_k}{Y_k^2}\right)^{-\frac{n+3}{2}}}{2^{n+1}\Gamma(n+1)} \cdot \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}n+\frac{3}{2})}{m!\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m \\ & \cdot F\left(-m, -m; n+1; 2\xi_k^2 Y_k^2\right) \end{aligned} \tag{3.9c}$$

$$C_k = \frac{Y_k^2}{2\gamma_k} \sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m \cdot F\left(-m, -m; 1; 2\xi_k^2 Y_k^2\right) \tag{3.9d}$$

$$D_k = \frac{8}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{\pi n}{4}\frac{\left(\frac{2\gamma_k}{Y_k}\right)^n\left(\frac{\gamma_k}{Y_k^2}\right)^{-\frac{n}{2}-1}}{2^{n+1}\Gamma(n+1)}\cdot\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}n+1)}{m!\Gamma(m+1)}\left(\frac{\gamma_k}{2\xi_k^2Y_k^2}\right)^m \quad (3.9e)$$
$$\cdot F\left(-m,-m;n+1;2\xi_k^2Y_k^2\right)$$

Equation Eq. (3.9a) gives the approximate Laplacian MMSE estimator of the spectral magnitudes. We will be referring to this estimator as the ApLapMMSE estimator.

3.4 Discussion on Convergence

Equation (3.6) and Eq. (3.9b)-Eq. (3.9e) involve infinite summations. Hence, their convergence is at question.

First, let us consider the convergence of Eq. (3.6). We found that it converges to the true value with 40 terms provided that $X_k \in [0, 10]$. For X_k taking larger values, however, we found that a larger number of terms is required, as demonstrated in Figures 3.6 and 3.7, which show divergence when X_k is evaluated at $X_k > 18$ but convergence again when the first 80 terms are evaluated. Hence, the larger X_k is, the more terms are needed in Eq. (3.6).

Next, we consider convergence of Eq. (3.9b)-Eq. (3.9e). We plot, in Figures 3.8-3.11, the first 40 elements of the four terms, A_k, B_k, C_k, and D_k, respectively, with $\xi_k = 1$, $\gamma_k = 1$, and $Y_k = 1$. For the second summation in terms B_k and D_k, we still use the first 40 terms. Note that the Bessel functions involved in the derived closed form solution may still cause numerical issues in practice. The solution, however, was

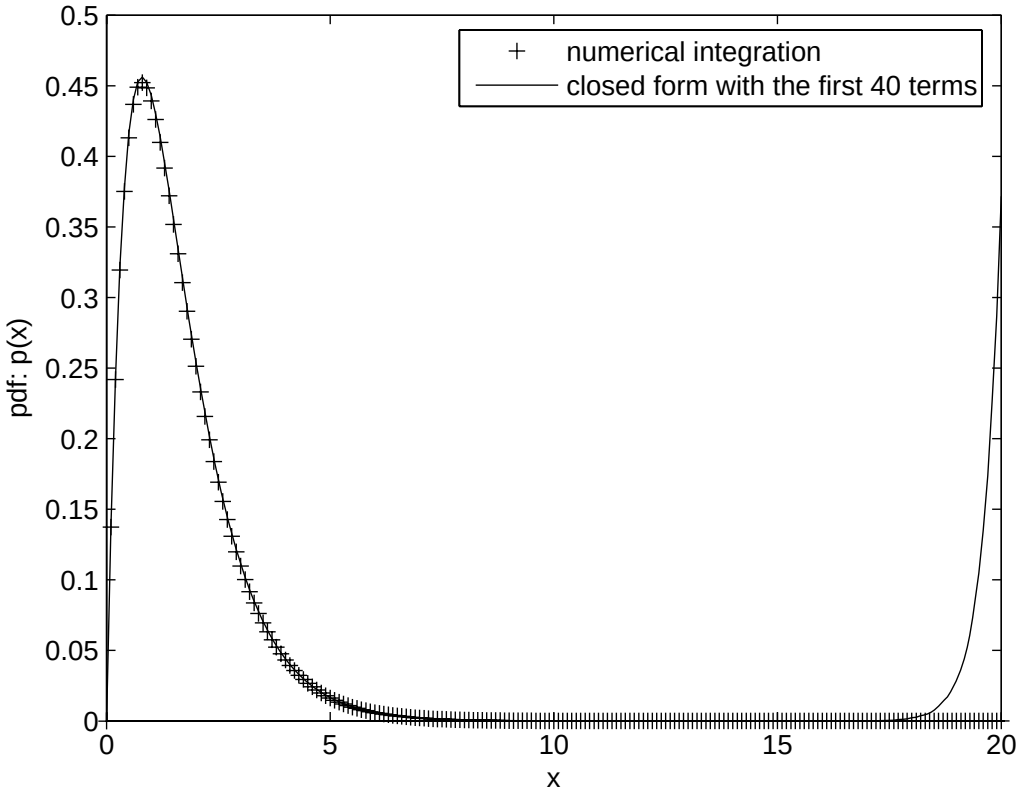

Figure 3.6. The solution diverges when magnitude is greater than 20 with the first 40 terms.

verified to converge empirically and theoretically.

As shown in Figures 3.8-3.11, the elements of all the four terms eventually converge to zero. Therefore, summations of those elements are bounded, and the solution given by Eq. (3.9b)-Eq. (3.9e) converges.

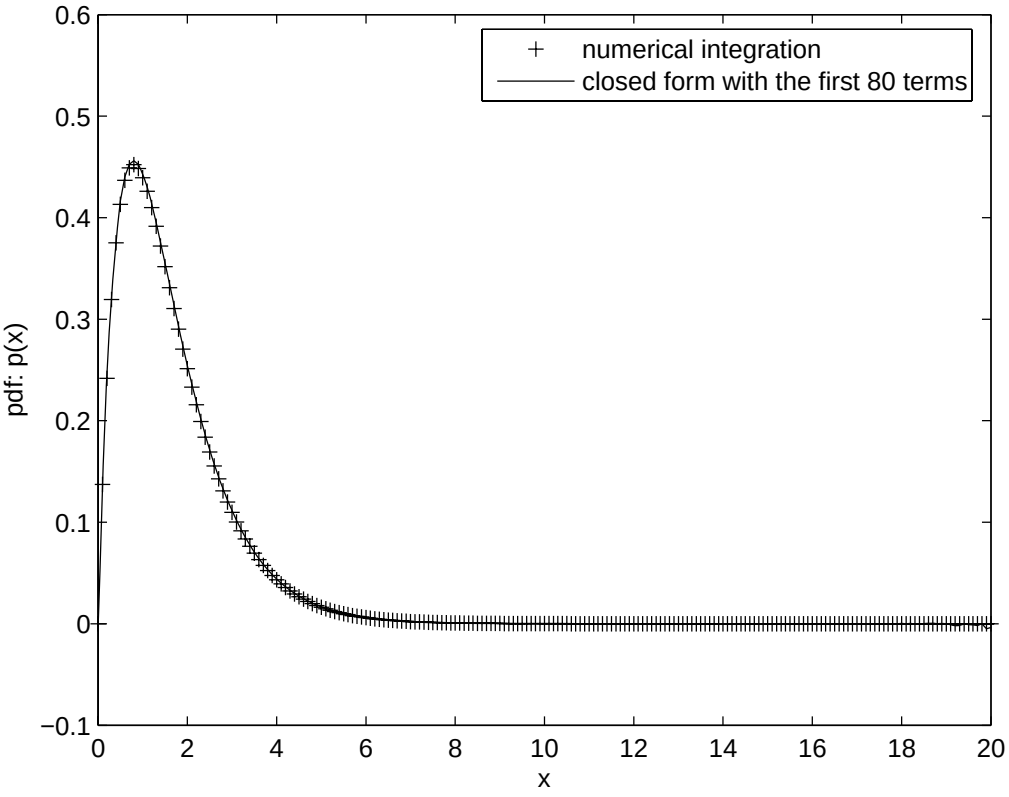

Figure 3.7. The solution converges when more (80) terms are evaluated.

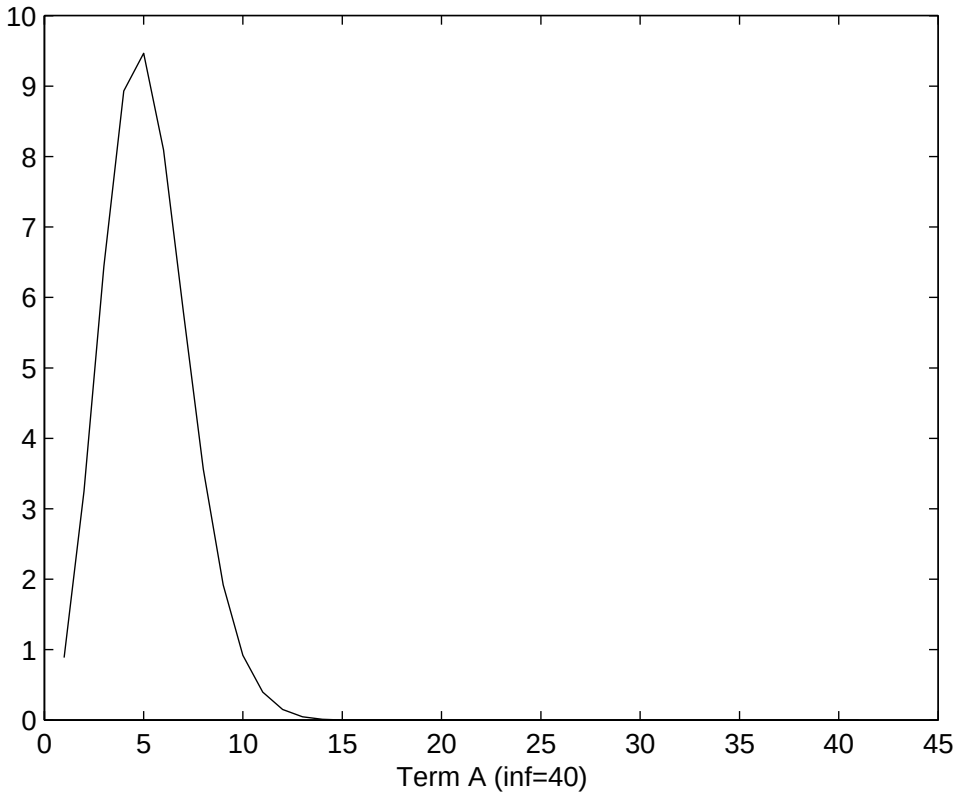

Figure 3.8. Plot of the first 40 elements in term A of Eq. (3.9b).

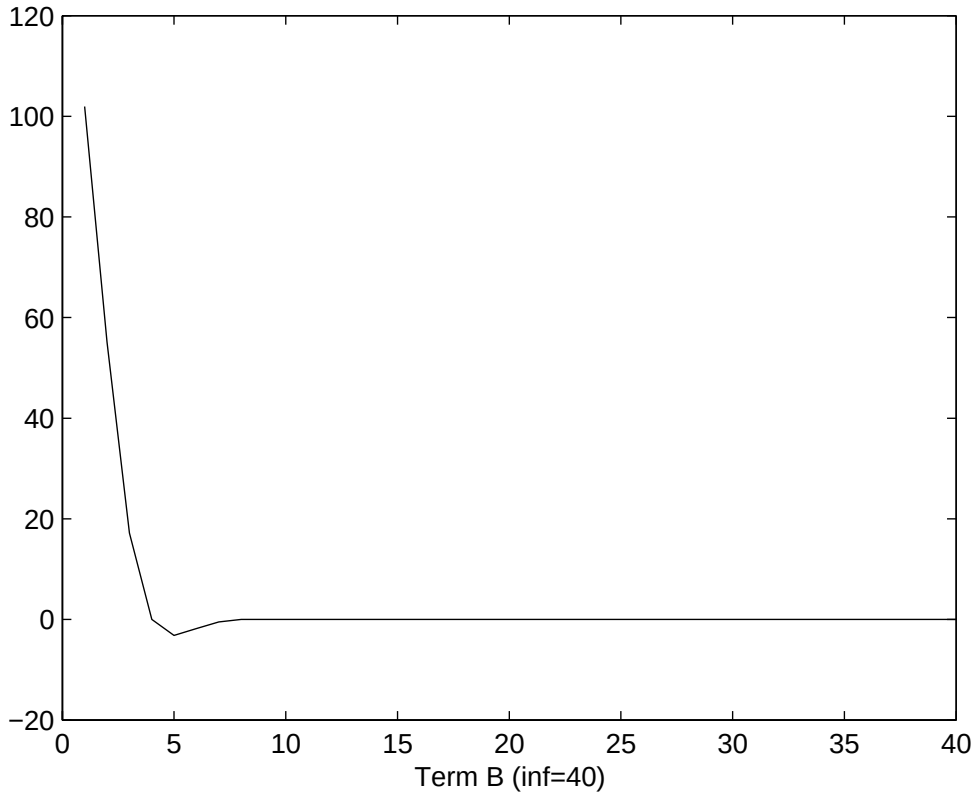

Figure 3.9. Plot of the first 40 elements in term B of Eq. (3.9c).

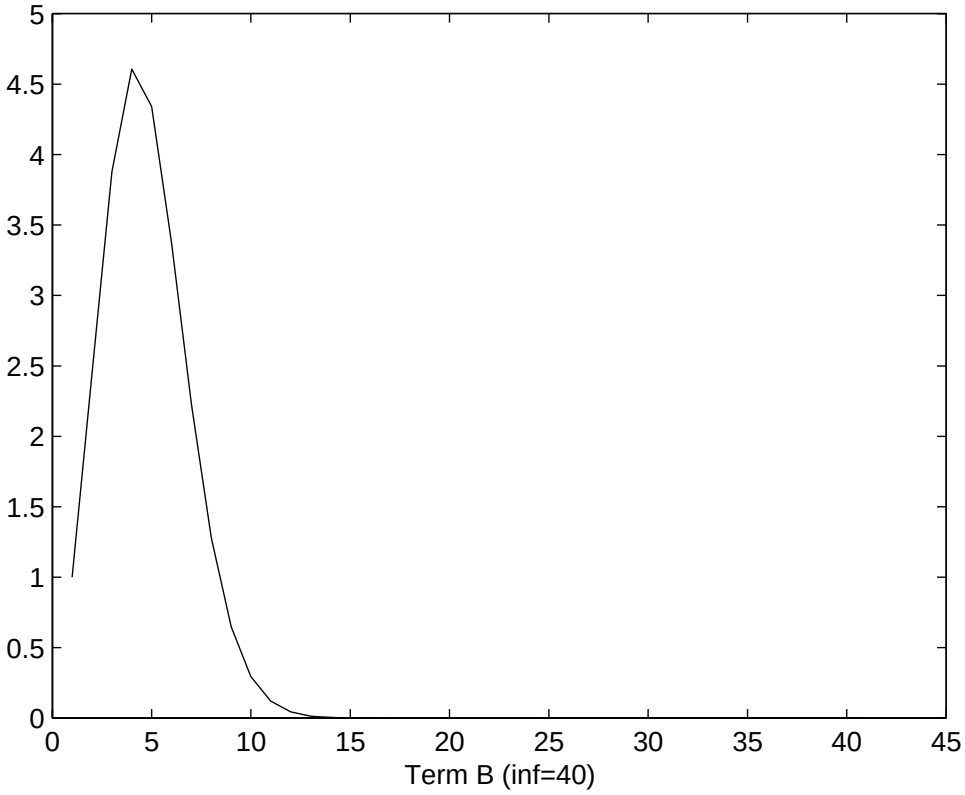

Figure 3.10. Plot of the first 40 elements in term C of Eq. (3.9d).

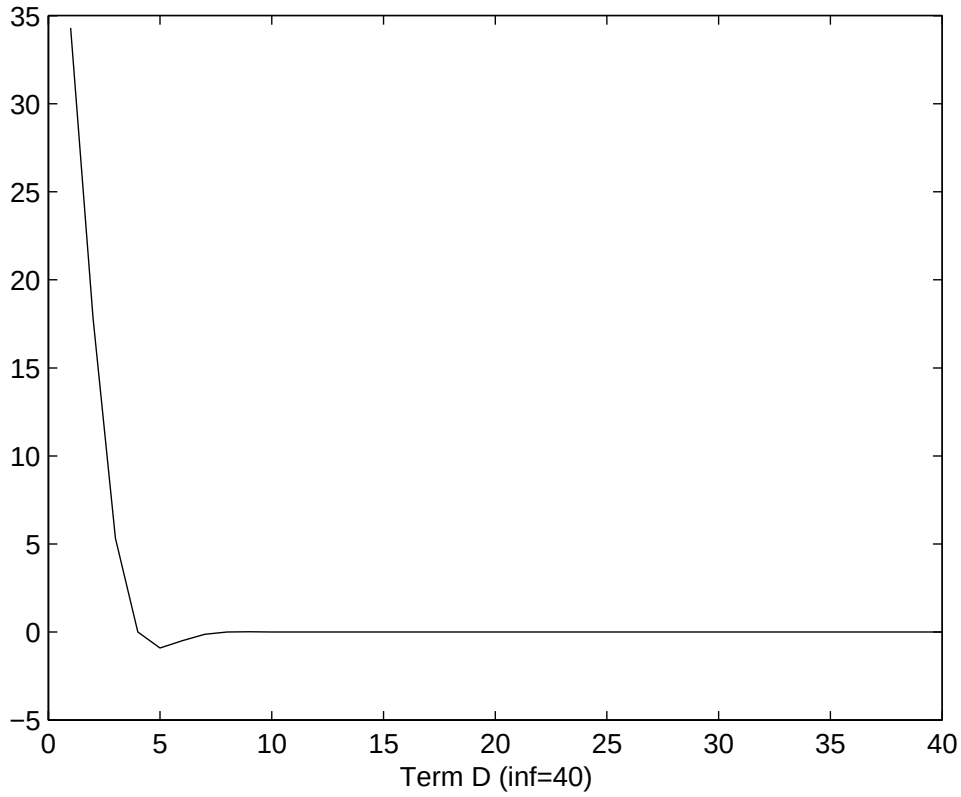

Figure 3.11. Plot of the first 40 elements in term D of Eq. (3.9e).

CHAPTER 4

AMPLITUDE ESTIMATOR UNDER SPEECH PRESENCE UNCERTAINTY

In the previous chapter, it was implicitly assumed that speech was present at all times. However, running speech contains a great deal of pauses, even during speech activity. The stop closures, for example, which are brief silence periods occurring before the burst of stop consonants, often appear in the middle of a sentence. Also, speech might not be present at a particular frequency even during voiced speech segments. In fact, this was exploited in multi-band speech coders where the spectrum was divided into bands, and each band was declared as being voiced or unvoiced (random-like). The voiced bands were assumed to be generated by a periodic excitation while the unvoiced bands were assumed to be generated by random noise. Such mixed-source excitation models were shown to produce better speech quality than the traditional voiced/unvoiced models. It follows then, that a better noise suppression rule might be produced if we assume a two-state model for speech events, that is, that either speech is present (at a particular frequency bin) or it is not.

Next, we present methods that incorporate the fact that speech might not be present at all frequencies and at all times. Intuitively, this amounts to multiplying the estimators Eq. (3.1) by a term that provides an estimate of the probability that speech is absent at a particular frequency bin. We call such estimators Speech-Presence

Uncertainty (SPU) estimators. SPU estimators are not new and have been studied in [46] [1] [47] [48]. None of these estimators incorporated Laplacian modeling of the DFT coefficients.

4.1 Incorporating Speech-Presence Uncertainty in MMSE Estimators

In this section, we derive the MMSE magnitude estimator under the assumed Laplacian model and uncertainty of speech presence. We consider a two-state model for speech events, that is, that either speech is present at a particular frequency bin (hypothesis H_1) or that is not (hypothesis H_0). This is expressed mathematically using the following binary hypothesis model:

$$H_0^k : \text{ speech absence: } Y(\omega_k) = D(\omega_k) \tag{4.1a}$$

$$H_1^k : \text{ speech present: } Y(\omega_k) = X(\omega_k) + D(\omega_k) \tag{4.1b}$$

To incorporate the above binary model to an MMSE estimator, we can use a weighted average of two estimators: one that is weighted by the probability that speech is present, and one that is weighted by the probability that speech is absent. So, if the original MMSE estimator had the form $\hat{X}_k = E(X_k|Y(\omega_k))$, then the new estimator has the form:

$$\hat{X}_k = E(X_k|Y(\omega_k), H_1^k)P(H_1^k|Y(\omega_k)) + E(X_k|Y(\omega_k), H_0^k)P(H_0^k|Y(\omega_k)) \tag{4.2}$$

where $P(H_1^k|Y(\omega_k))$ denotes the conditional probability that speech is present in frequency bin k given the noisy speech spectrum. Similarly, $P(H_0^k|Y(\omega_k))$ denotes the

conditional probability that speech is absent given the noisy speech spectrum. The term $E(X_k|Y(\omega_k), H_0^k)$ in the above equation is zero since it represents the average value of X_k given the noisy spectrum $Y(\omega_k)$ and the fact that speech is absent. Therefore, the MMSE estimator in Eq. (4.2) reduces to:

$$\hat{X}_k = E(X_k|Y(\omega_k), H_1^k)P(H_1^k|Y(\omega_k)) \tag{4.3}$$

Bayes' rule [43] can be used to compute $P(H_1^k|Y(\omega_k))$. The MMSE estimator of the spectral component at frequency bin k is weighted by the probability that speech is present at that frequency:

$$\begin{aligned} P(H_1^k|Y(\omega_k)) &= \frac{p(Y(\omega_k)|H_1^k)P(H_1)}{p(Y(\omega_k)|H_1^k)P(H_1^k) + p(Y(\omega_k)|H_0^k)P(H_0)} \\ &= \frac{\Lambda(Y(\omega_k), q_k)}{1 + \Lambda(Y(\omega_k), q_k)} \end{aligned} \tag{4.4}$$

where $\Lambda(Y(\omega_k), q_k)$ is the generalized likelihood ratio defined by:

$$\Lambda(Y(\omega_k), q_k) = \frac{1 - q_k}{q_k} \frac{p(Y(\omega_k)|H_1^k)}{p(Y(\omega_k)|H_0^k)} \tag{4.5}$$

where $q_k = P(H_0^k)$ denotes the *a priori* probability of speech absence for frequency bin k. The a priori probability of speech presence, i.e., $P(H_1^k)$, is given by $(1 - q_k)$. Theoretically, the optimal estimate under hypothesis H_0^k is identical to zero but a small nonzero value might be preferable for perceptual purposes [49].

Under hypothesis H_0, $Y(\omega_k) = D(\omega_k)$, and given that the noise is complex Gaussian with zero mean and variance $\lambda_d(k)$, it follows that $p(Y(\omega_k)|H_0^k)$ will also have

a Gaussian distribution with the same variance, i.e.,

$$p(Y(\omega_k)|H_0^k) = \frac{1}{\pi\lambda_d(k)} \exp\left(-\frac{Y_k^2}{\lambda_d(k)}\right) \tag{4.6}$$

If $X(\omega_k)$ follows a Laplacian distribution, we need to compute $p(Y(\omega_k)|H_1^k)$. Assuming independence between real and imaginary components, we have:

$$p(Y(\omega_k)|H_1^k) = p(z_r, z_i) = p_{Z_r(k)}(z_r)p_{Z_i(k)}(z_i) \tag{4.7}$$

where $Z_r(k) = Re\{Y(\omega_k)\}$, and $Z_i(k) = Im\{Y(\omega_k)\}$.

Under hypothesis H_1, we need to derive the pdf of $Y(\omega_k) = X(\omega_k) + D(\omega_k)$, where $X(\omega_k) = X_r(\omega_k) + jX_i(\omega_k)$ and $D(\omega_k) = D_r(\omega_k) + jD_i(\omega_k)$. The pdfs of $X_r(\omega_k)$ and $X_i(\omega_k)$ are assumed to be Laplacian and the pdfs of $D_r(\omega_k)$ and $D_r(\omega_k)$ are assumed to be Gaussian with variance $\sigma_d^2/2$ and zero mean. We assume that $p_{X_r}(x_r) = \frac{1}{2\sigma_x}\exp(-\frac{|x_r|}{\sigma_x})$, and $p_{D_r}(d_r) = \frac{1}{\sqrt{\pi}\sigma_d}\exp(-\frac{d_r^2}{\sigma_d^2})$. The derivation of 4.7 is given in Appendix C. The solution for Eq. (4.7) is given by:

$$\begin{aligned} p_{Z_r(k)}(z_r) &= \frac{\sqrt{\gamma_k}\exp(\frac{1}{2\xi_k})}{2\sqrt{2\xi_k}Y_k}\left[\exp(-\frac{\sqrt{\gamma_k}z_r}{Y_k\sqrt{\xi_k}}) + \exp(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k}) + \right. \\ &\left. \exp(-\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k} - \frac{1}{\sqrt{\xi_k}}) - \exp(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k} + \frac{1}{\sqrt{\xi_k}})\right] \end{aligned} \tag{4.8}$$

$$\begin{aligned} p_{z_i(k)}(z_i) &= \frac{\sqrt{\gamma_k}\exp(\frac{1}{2\xi_k})}{2\sqrt{2\xi_k}Y_k}\left[\exp(-\frac{\sqrt{\gamma_k}z_i}{Y_k\sqrt{\xi_k}}) + \exp(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k}) + \right. \\ &\left. \exp(-\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k} - \frac{1}{\sqrt{\xi_k}}) - \exp(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k} + \frac{1}{\sqrt{\xi_k}})\right] \end{aligned} \tag{4.9}$$

where $erf(.)$ is the error function [44], $\xi_k = \lambda_x(k)/\lambda_d(k)$ (*a priori* SNR) and $\gamma_k = Y_k^2/\lambda_d(k)$ (*a posteriori* SNR). Simply substituting Eq. (4.7) and Eq. (4.6) into Eq. (4.4) and Eq. (4.5), we obtain the Speech Presence Uncertainty (SPU) estimator.

4.2 Relation between SPU and *a priori* SNR ξ_k

Intuitively, the relation between SPU and *a priori* SNR ξ_k should remain the same whether the speech signal is modeled by Gaussian or Laplacian distribution. In order to illustrate this relationship, we examined it under the condition that the speech signal is modeled by a Gaussian pdf. In this case, the pdf $p(Y(\omega_k)|H_1^k)$ is given by:

$$p(Y(\omega_k)|H_1^k) = \frac{1}{\pi\left[\lambda_d(k) + \lambda_x(k)\right]} \exp\left(-\frac{Y_k^2}{\lambda_d(k) + \lambda_x(k)}\right) \tag{4.10}$$

Note that under hypothesis H_1, $Y(\omega_k) = X(\omega_k) + D(\omega_k)$, and the pdfs of $X(\omega_k)$ and $D(\omega_k)$ are complex Gaussian with zero mean and variances $\lambda_d(k)$ and $\lambda_x(k)$ respectively. It follows that $Y(\omega_k)$ will also have a Gaussian distribution with variance $\lambda_d(k)+\lambda_x(k)$. The expression for the likelihood ratio is given by Eq. (4.5):

$$\Lambda(Y(\omega_k), q_k, \xi_k') = \frac{1-q_k}{q_k} \frac{\exp\left[\frac{\xi_k'}{1+\xi_k'}\gamma_k\right]}{1+\xi_k'} \tag{4.11}$$

where ξ_k' indicates the conditional *a priori* SNR:

$$\xi_k' \triangleq \frac{E\left[X_k^2|H_1^k\right]}{\lambda_d(k)} \tag{4.12}$$

Note that ξ_k is the unconditional *a priori* SNR, whose relation with the conditional *a priori* SNR defined in Eq. (4.12) can be expressed as follows

$$\begin{aligned} \xi_k &= \frac{E\left[X_k^2\right]}{\lambda_d(k)} = p(H_1^k)\frac{E\left[X_k^2|H_1^k\right]}{\lambda_d(k)} \\ &= (1-q_k)\xi_k' \end{aligned} \tag{4.13}$$

Substituting Eq. (4.11) into Eq. (4.4), the *a posteriori* probability of speech presence becomes:

$$P(H_1^k|Y(\omega_k)) = \frac{1-q_k}{1-q_k+q_k(1+\xi_k')\exp(-\upsilon_k')} \tag{4.14}$$

where $\upsilon'_k = \frac{\xi'_k}{\xi'_k+1}\gamma_k$

From Eq. (4.14) we note that when ξ'_k is large, speech is surely present, and $P(H_1^k|Y(\omega_k)) \approx 1$, as expected. On the other hand, when ξ'_k is extremely small, $P(H_1^k|Y(\omega_k)) \approx 1 - q_k$, i.e., it is equal to the a priori probability of speech presence, $P(H_1^k)$.

For Laplacian modeling, it is very difficult to derive an explicit relation between ξ_k and $P(H_1^k|Y(\omega_k))$ as in Eq. (4.14), since the $P(H_1^k|Y(\omega_k))$ term derived with Laplacian modeling is highly non-linear. Therefore, we plotted $P(H_1^k|Y(\omega_k))$ to facilitate our comparison with the Gaussian-based $P(H_1^k|Y(\omega_k))$ (Eq. (4.14)). Figure 4.1 plots $P(H_1^k|Y(\omega_k))$ as a function of a *priori* SNR ξ_k for different values of the noise variance λ_d and for $q_k = 0.3$ (we did not make use of ξ'_k in the derived estimator). As the Figure shows, both estimates of $P(H_1^k|Y(\omega_k))$ follow the same trend. As ξ_k increases, $P(H_1^k|Y(\omega_k)) \longrightarrow 1$. The main difference lies in extremely low values of ξ_k. The estimate of $P(H_1^k|Y(\omega_k))$ based on Laplacian modeling progressively goes to zero as $\xi_k \longrightarrow 0$, while the estimate of $P(H_1^k|Y(\omega_k))$ based on Gaussian modeling (Eq. (4.14)) approaches $P(H_1^k|Y(\omega_k)) \approx 1 - q_k$.

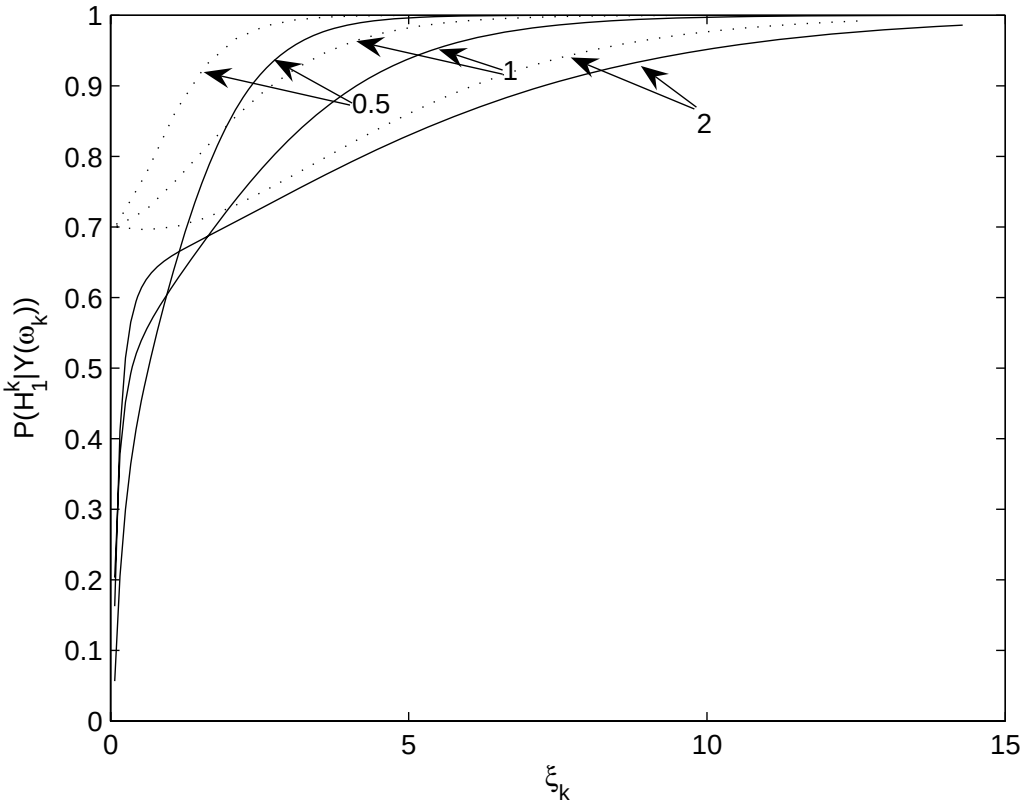

Figure 4.1. $P(H_1^k|Y(\omega_k))$ as a function of a *priori* SNR ξ_k for the noise variance $\lambda_d = 0.5$, 1, and 2, respectively. The solid line plots $P(H_1^k|Y(\omega_k))$ for the Laplacian distribution, and the dotted line plots $P(H_1^k|Y(\omega_k))$ for the Gaussian distribution as in Eq. (4.14)

CHAPTER 5
IMPLEMENTATION AND PERFORMANCE EVALUATION

As discussed in Chapter 3, the closed form solution (Eq. (3.9a)) involved Bessel functions and infinite series. Numerical computation issues had to be taken into account in the implementation, and these are discussed next.

5.1 Numerical Integration

We have shown that retaining only the first 40 terms in Eq. (3.6) gave a good approximation of $p(x_k)$. This is demonstrated in Figure 3.1, which shows $p(x_k)$ estimated using numerical integration techniques and also approximated by truncating the summation in Eq. (3.6) to the first 40 terms. However, as also shown in Eq. (3.9a), the derived ApLapMMSE estimator is highly nonlinear since it involves infinite summations. We initially truncated the infinite summations to a large number of terms, however, simulations indicated that such an approximation led to numerical instability issues. That was due to the characteristics of the Bessel function diverging with spectral magnitudes increasing. For that reason, we chose to use numerical integration techniques [50] to evaluate the integrals in Eq. (3.1). Numerical integration techniques were also used to evaluate the integrals involved in the LapMMSE estimator in Eq. (3.4).

Many numerical integration techniques exist in the literature. For this work,

we choose the techniques in [50] due to its lower computation complexity than others. We extended the numerical integration technique to calculate the integral of product of two functions as shown in Eq. (5.1). More specifically, the following equation was used (see Appendix D):

$$\begin{aligned}\int_C^D g(x)F(x)dx &= \frac{\Delta}{3}g(x_0)F(x_0) + \frac{\Delta}{6}g(x_0)F(x_1) \\ &+\frac{\Delta}{3}g(x_M)F(x_M) + \frac{\Delta}{6}g(x_{M-1})F(x_M) \\ &+\sum_{l=1}^{M-1}\left[\frac{2\Delta}{3}g(x_l)F(x_l) + \frac{2\Delta}{3}g(x_{l+1})F(x_l) + \frac{\Delta}{6}g(x_l)F(x_{l-1})\right]\end{aligned} \tag{5.1}$$

where the interval $[C, D]$ is divided evenly into M subintervals and $\Delta = \frac{D-C}{M}$. Determination of M is based on a trade-off between accuracy and computation load. We applied the above numerical integration technique to find the solutions of both Eq. (3.4) and Eq. (B.2). MATLAB source code is provided in Appendix F.

In the simulations, we found that the terms B and D in Eq. (3.9a) are always very small compared to terms A and C. So those two terms were neglected from computation to further reduce the computational load.

5.2 Decision-Directed Estimation Approach

In the proposed estimator, the *a priori* SNR ξ_k is unknown and we have to estimate it in order to implement the estimator (Eq. (3.9a)). The reason ξ_k is unknown is because the clean signal is unavailable. The "decision-directed" approach [1] was used to estimate ξ_k.

As before, let $\xi_k(n)$, $X_k(n)$, $\lambda_d(k,n)$, and $\gamma_k(n)$ denote the *a priori* SNR, the

magnitude, the noise variance, and the *a posteriori* SNR, respectively, of the corresponding kth spectral component in the nth analysis frame. The derivation of the *a priori* SNR estimator is based on the definition of ξ_k, and its relation to the *a posteriori* SNR γ_k, as given below:

$$\xi_k(n) = \frac{E\{X_k^2(n)\}}{\lambda_d(k,n)} \tag{5.2}$$

$$\xi_k(n) = E\{\gamma_k(n) - 1\} \tag{5.3}$$

Combining Eq. (5.2) and Eq. (5.3) we have

$$\xi_k(n) = E\left\{\frac{1}{2}\frac{X_k^2(n)}{\lambda_d(k,n)} + \frac{1}{2}\left[\gamma_k(n) - 1\right]\right\} \tag{5.4}$$

The proposed estimator $\hat{\xi}_k$ of ξ_k is deduced from Eq. (5.4), and is given by:

$$\hat{\xi}_k(n) = \alpha\frac{\hat{X}_k^2(n-1)}{\lambda_d(k,n-1)} + (1-\alpha)P[\gamma_k(n) - 1],\ 0 \leqslant \alpha < 1 \tag{5.5}$$

where $\hat{X}_k(n-1)$ is the estimated amplitude of the kth signal spectral component in the $(n-1)$th analysis frame, and $P[\cdot]$ is an operator defined by

$$P[x] = \begin{cases} x & \text{if } x \geqslant 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

We see that $\hat{\xi}_k(n)$ is obtained from Eq. (5.4) by dropping the expectation operator, and using use the estimated amplitude of the $(n-1)$th frame instead of the amplitude in the current frame which we do not know. $P[\cdot]$ is used to ensure the positiveness of the proposed estimator in case $(\gamma_k(n) - 1)$ is negative.

The above estimator for $\xi_k(n)$ is a "decision-directed" type estimator, since $\hat{\xi}_k(n)$ is updated on the basis of a previous amplitude estimate. By using $X_k(n) = G(\hat{\xi}_k(n),$

$\gamma_k(n))R_k(n)$, where $G(\cdot,\cdot)$ is a gain function which results from the proposed amplitude estimator, Eq. (5.5) can be written in a way which emphasizes its recursive nature. From Eq. (5.5), we get

$$\hat{\xi}_k(n) = \alpha G^2(\hat{\xi}_k(n-1),\, \gamma_k(n-1))\gamma_k(n-1) + (1-\alpha)P[\gamma_k(n)-1] \qquad (5.7)$$

The initial conditions need to be determined and $\xi_k(0) = \alpha + (1-\alpha)P[\gamma_k(0)-1]$ is found appropriate since it minimizes initial transition effects in the enhanced speech.

The theoretical investigation of the recursive estimator Eq. (5.7) is very complicated due to its highly nonlinear nature, even for the simple gain function of the Wiener amplitude estimator. We used $\alpha = 0.98$ in Eq. (5.7) for all simulations.

5.3 Performance Evaluation

Twenty sentences from the TIMIT database were used for the objective evaluation of the proposed LapMMSE estimator, 10 produced by female speakers and 10 produced by male speakers. The TIMIT sentences were downsampled to 8 kHz. Speech-shaped noise constructed from the long-term spectrum of the TIMIT sentences as well as F-16 cockpit noise were added to the clean speech files at 0, 5 and 10 dB SNR. An estimate of the noise spectrum was obtained from the initial 100-ms segment of each sentence [28] [51]. The noise spectrum estimate was not updated in subsequent frames.

The proposed estimators were applied to 20-ms duration frames of speech using a Hamming window [52], with 50% overlap between frames. The enhanced signal was combined using the overlap and add approach [53] [54]. The STSA of the speech

signal was then estimated, and combined with the complex exponential of the noisy phase. The "decision-directed" approach [1] introduced above was used in the proposed estimators to compute the *a priori* SNR ξ_k, with $\alpha = 0.98$. The *a priori* probability of speech absence, q_k, was set to $q_k = 0.3$ in Eq. (4.5). q_k could be estimated as a function of frequency and time [47] but we empirically determined it for simplicity here.

Figures 5.1 and 5.2 show spectrograms of the signal in quiet, signal in speech-shaped noise, signal enhanced by the Gaussian MMSE estimator, signal enhanced by the LapMMSE estimator and signal enhanced by the ApLapMMSE estimator. Similarly, Figures 5.3 and 5.4 show spectrograms of the signal in quiet, signal in F-16 cockpit noise, signal enhanced by the Gaussian MMSE estimator, signal enhanced by the LapMMSE estimator and signal enhanced by the ApLapMMSE estimator. It is clear that the proposed Laplacian-based estimators yielded less residual noise than the Gaussian-based estimator, and enhanced speech produced by the LapMMSE and ApLapMMSE estimators looked very similar.

Objective measures [55] were used to evaluate the performance of the proposed estimators implemented with and without speech presence uncertainty (SPU) and denoted as LapMMSE-SPU and LapMMSE respectively. Similarly, the approximate Laplacian estimators implemented with and without speech presence uncertainty were indicated as ApLapMMSE-SPU and ApLapMMSE respectively. For comparative purposes we evaluated the performance of the traditional (Gaussian-based) MMSE estimator [1] with and without incorporating speech presence uncertainty which were indicated as MMSE-SPU and MMSE respectively. We also evaluated the performance

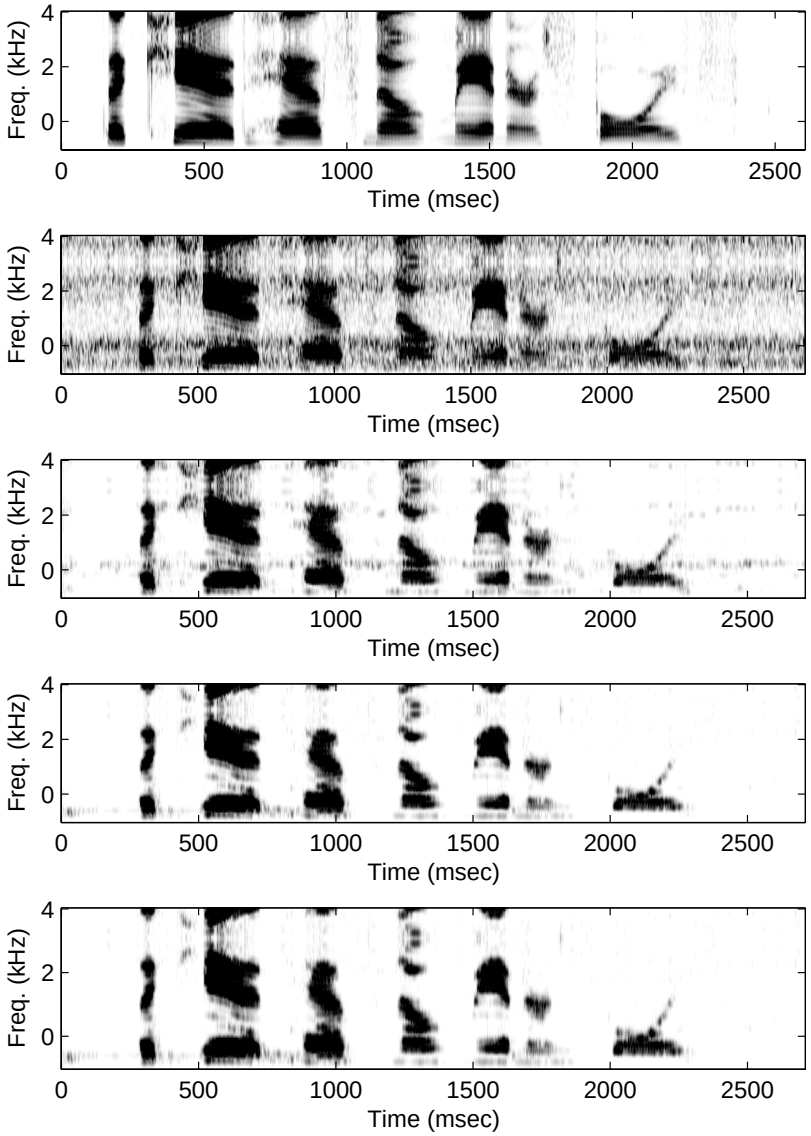

Figure 5.1. Spectrograms of a TIMIT sentence spoken by a female speaker and enhanced by the Gaussian and Laplacian MMSE estimators. From top to bottom, are the spectrograms of the signal in quiet, signal in speech-shaped noise, signal enhanced by the Gaussian MMSE estimator [1], signal enhanced by the LapMMSE estimator and signal enhanced by the ApLapMMSE estimator. All estimators incorporated signal-presence uncertainty.

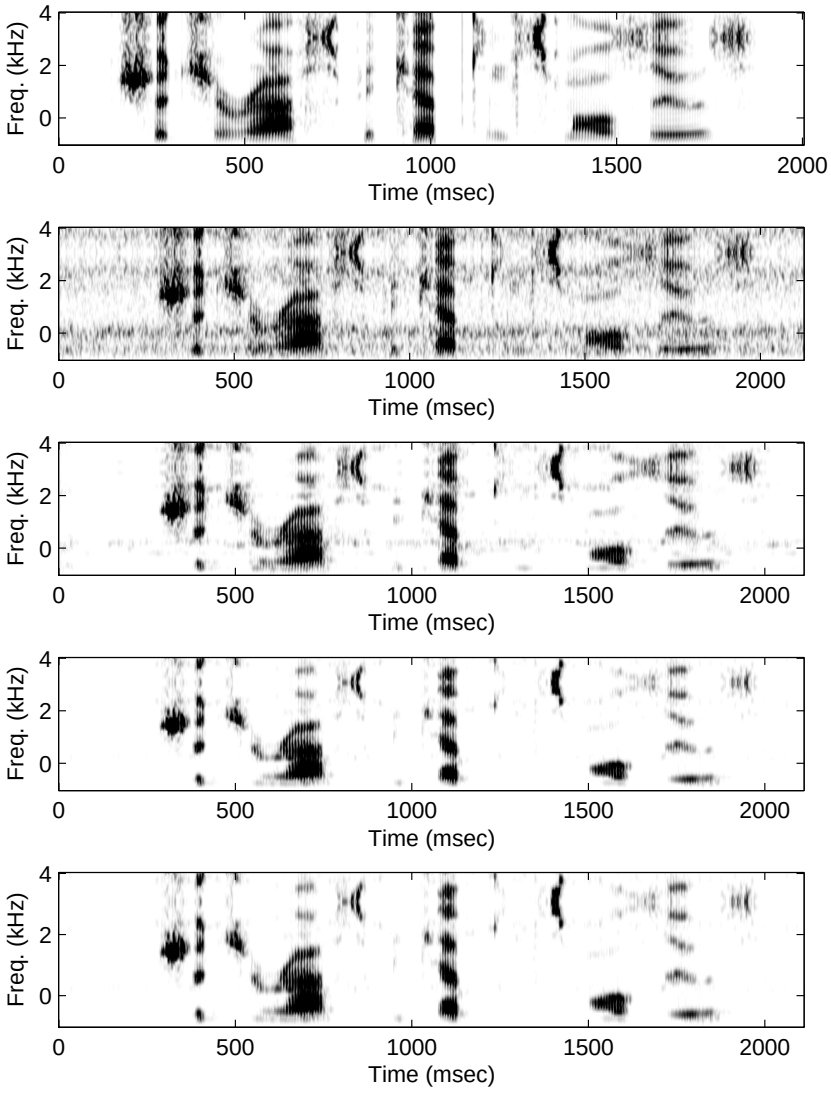

Figure 5.2. Spectrograms of a TIMIT sentence spoken by a male speaker and enhanced by the Gaussian and Laplacian MMSE estimators. From top to bottom, are the spectrograms of the signal in quiet, signal in speech-shaped noise, signal enhanced by the Gaussian MMSE estimator [1], signal enhanced by the LapMMSE estimator and signal enhanced by the ApLapMMSE estimator. All estimators incorporated signal-presence uncertainty.

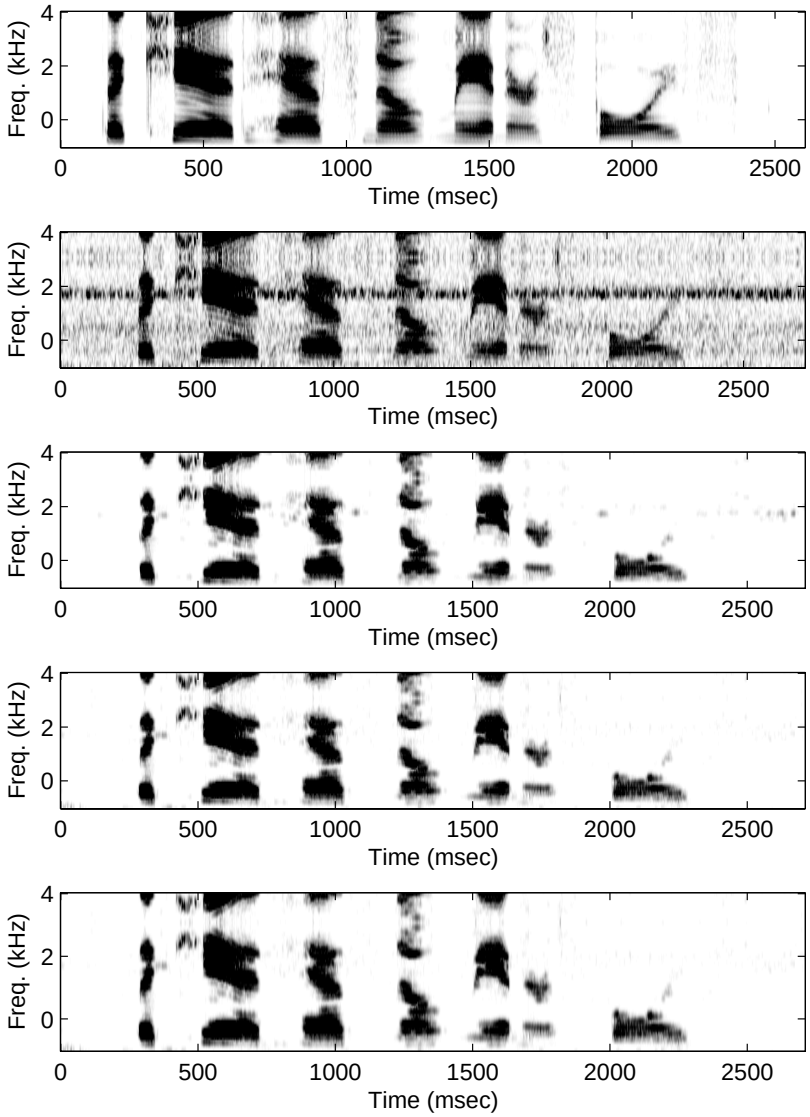

Figure 5.3. Spectrograms of a TIMIT sentence spoken by a female speaker and enhanced by the Gaussian and Laplacian MMSE estimators. From top to bottom, are the spectrograms of the signal in quiet, signal in F-16 noise, signal enhanced by the Gaussian MMSE estimator [1], signal enhanced by the LapMMSE estimator and signal enhanced by the ApLapMMSE estimator. All estimators incorporated signal-presence uncertainty.

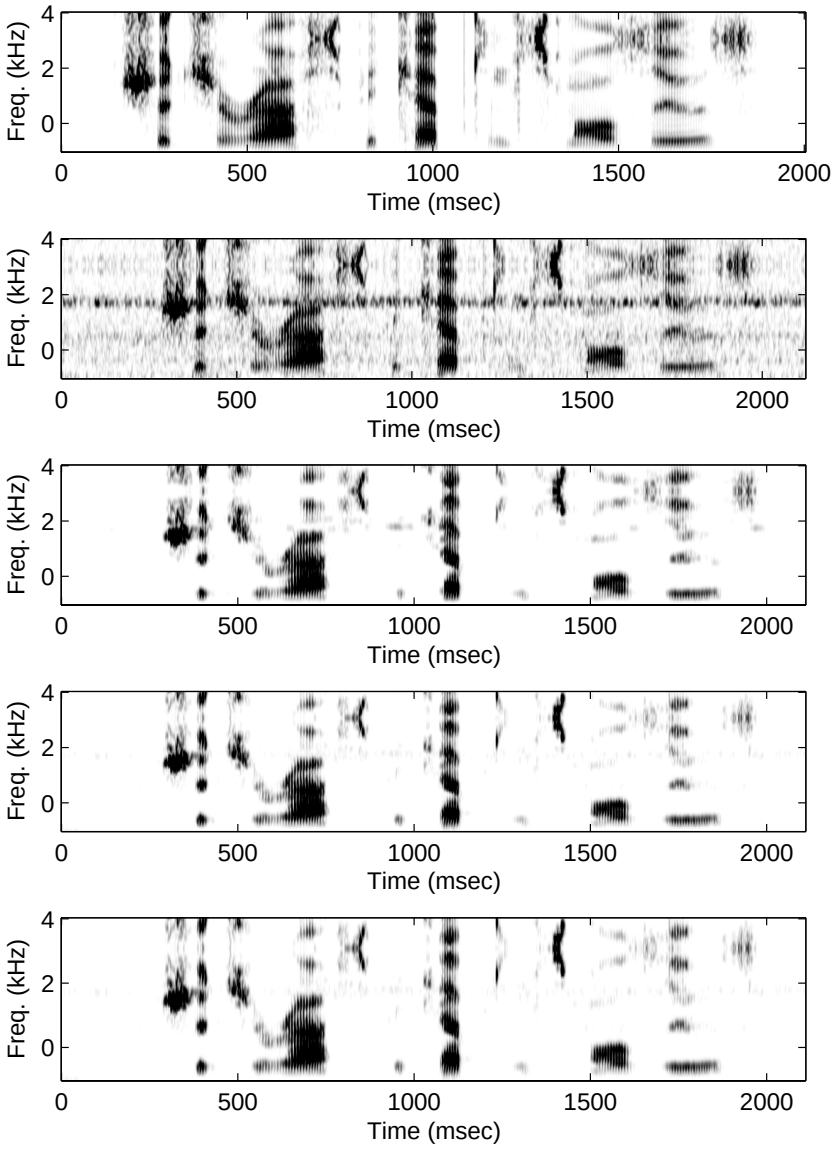

Figure 5.4. Spectrograms of a TIMIT sentence spoken by a male speaker and enhanced by the Gaussian and Laplacian MMSE estimators. From top to bottom, are the spectrograms of the signal in quiet, signal in F-16 noise, signal enhanced by the Gaussian MMSE estimator [1], signal enhanced by the LapMMSE estimator and signal enhanced by the ApLapMMSE estimator. All estimators incorporated signal-presence uncertainty.

of the complex-valued MMSE estimator derived in [35] based on Laplacian speech priors. Note that in [35], the estimator $E[X(\omega_k)|Y(\omega_k)]$ was derived by combining the estimators of the real and imaginary parts of the DFT coefficients.

The segmental SNR, log likelihood ratio and PESQ (ITU-T P.862) [56] [57] measures were used for objective evaluation of the proposed estimators. The segmental SNR was computed as (Appendix D):

$$SNRseg = \frac{10}{M} \sum_{k=0}^{M-1} \log_{10} \frac{\sum_{n=Nk}^{Nk+N-1} x^2(n)}{\sum_{n=Nk}^{Nk+N-1} \left(x(n) - \widehat{x}(n)\right)^2} \tag{5.8}$$

where M is the total number of frames, N is the frame size, $x(n)$ is the clean signal and $\widehat{x}(n)$ is the enhanced signal. Since the segmental SNR can become very small and negative during periods of silence, we limited the $SNRseg$ values to the range of [-10 dB, 35 dB] as per [58]. In doing so, we avoided the explicit marking and identification of speech-absent segments.

The log likelihood ratio (LLR) [59] for each 20-ms frame was computed as (Appendix D) :

$$LLR = \log \left(\frac{a_d^T R_x a_d}{a_x^T R_x a_x} \right) \tag{5.9}$$

where a_x and a_d are the prediction coefficients of the clean and enhanced signals respectively, and R_x is the autocorrelation matrix of the clean signal. The mean LLR value was computed across all frames for each sentence. Since the mean can be easily biased by a few outlier frames, we computed the mean based on the lowest 95% of the frames as per [58].

Enhanced speech was also evaluated using the PESQ algorithm [56]. The PESQ algorithm is designed to predict subjective opinion scores of degraded speech samples. PESQ returns a score from 4.5 to -0.5, with higher scores indicating better quality.

Tables 5.1 and 5.2 list the segmental SNR values and log-likelihood ratio values obtained by the various estimators at different SNRs, and Table 5.3 lists the PESQ values. As can be seen, higher segmental SNR values and higher PESQ values were obtained consistently by the proposed estimators (LapMMSE and ApLapMMSE). Particularly large improvements were noted at higher SNR levels (5 and 10 dB). The segmental SNR values obtained by the proposed Laplacian-based MMSE estimators were 3-5 dB higher than the segmental SNR values obtained by the Gaussian-based MMSE estimator and 2-3 dB higher than the segmental SNR obtained by the Laplacian estimator proposed in [35]. The Laplacian estimator proposed in [35] also performed better than the Gaussian MMSE estimator. The difference in performance between the LapMMSE and ApLapMMSE estimators was extremely small, suggesting that our assumptions about the independence of magnitude and phase were reasonable and did not cause any significant performance degradation.

The pattern of results was similar with the log-likelihood ratio objective measure (Table 5.2). Smaller LLR values were obtained by the proposed Laplacian estimators compared to the Gaussian-based MMSE estimator for all SNR conditions. This was consistent with the segmental SNR objective evaluation (Table 5.1).

Informal listening tests indicated that speech enhanced by the Laplacian MMSE estimators had less residual noise. This was confirmed by visual inspection of spec-

trograms of the enhanced speech signals. Figure 5.1 shows the spectrograms of the TIMIT sentence “The kid has no manners, boys” enhanced by the LapMMSE-SPU, ApLapMMSE-SPU and MMSE-SPU estimators. Similarly, Figure 5.2 shows the same results from the TIMIT sentence "Should giraffes be kept in small zoos?". The sentence was originally embedded in +5 dB S/N speech-shaped noise. As can be seen, the sentence enhanced by the Laplacian MMSE estimators had less residual noise with no compromise in speech distortion. The quality of speech enhanced by the ApLapMMSE and LapMMSE estimators was nearly identical, consistent with the objective evaluation of these estimators (Tables 5.1-5.3). The data in Tables 5.1-5.3 are also plotted in Figures 5.5-5.7. Note that only results for the ApLapMMSE estimator are plotted in those figures for better clarity since the performance of the LapMMSE and ApLapMMSE was similar.

We plotted the spectrograms of the same sentences corrupted by F-16 noise for comparative purposes in Figures 5.3 and 5.4. As expected, the enhanced results in F-16 noise are better than those in speech-shaped noise probably due to the spectral similarity of the spectrum of the speech-shaped noise with the speech spectrum. The proposed LapMMSE estimator produced less residual noise than the Gaussian MMSE estimator although the difference is less obvious compared to the speech-shaped noise.

Estimator	Speech-Shaped			F-16 Fighter		
	0dB	5dB	10dB	0dB	5dB	10dB
MMSE	0.763	1.96	2.979	1.414	2.285	3.048
LapMMSE	1.202	4.414	7.301	1.915	5.159	7.533
ApLapMMSE	1.149	4.647	7.182	1.819	5.122	7.528
MMSE-Lap [35]	1.773	2.652	3.594	2.963	4.414	6.348
MMSE-SPU	0.859	2.027	3.067	1.495	2.362	3.125
LapMMSE-SPU	1.606	4.931	7.561	2.775	5.804	8.356
ApLapMMSE-SPU	1.867	5.113	7.792	2.650	5.657	8.198

Table 5.1. Comparative performance, in terms of segmental SNR, of the Gaussian-based MMSE and Laplacian-based MMSE estimators.

Estimator	Speech-Shaped			F-16 Fighter		
	0dB	5dB	10dB	0dB	5dB	10dB
MMSE	1.433	1.11	0.852	1.094	0.86	0.676
LapMMSE	0.789	0.615	0.589	0.602	0.477	0.467
ApLapMMSE	0.783	0.618	0.603	0.598	0.479	0.479
MMSE-Lap [35]	1.017	0.772	0.610	0.870	0.693	0.554
MMSE-SPU	1.113	0.754	0.625	0.85	0.584	0.496
LapMMSE-SPU	0.756	0.598	0.584	0.577	0.464	0.463
ApLapMMSE-SPU	0.763	0.602	0.591	0.582	0.467	0.469

Table 5.2. Comparative performance, in terms of log-likelihood ratio, of the Gaussian-based MMSE and Laplacian-based MMSE estimators.

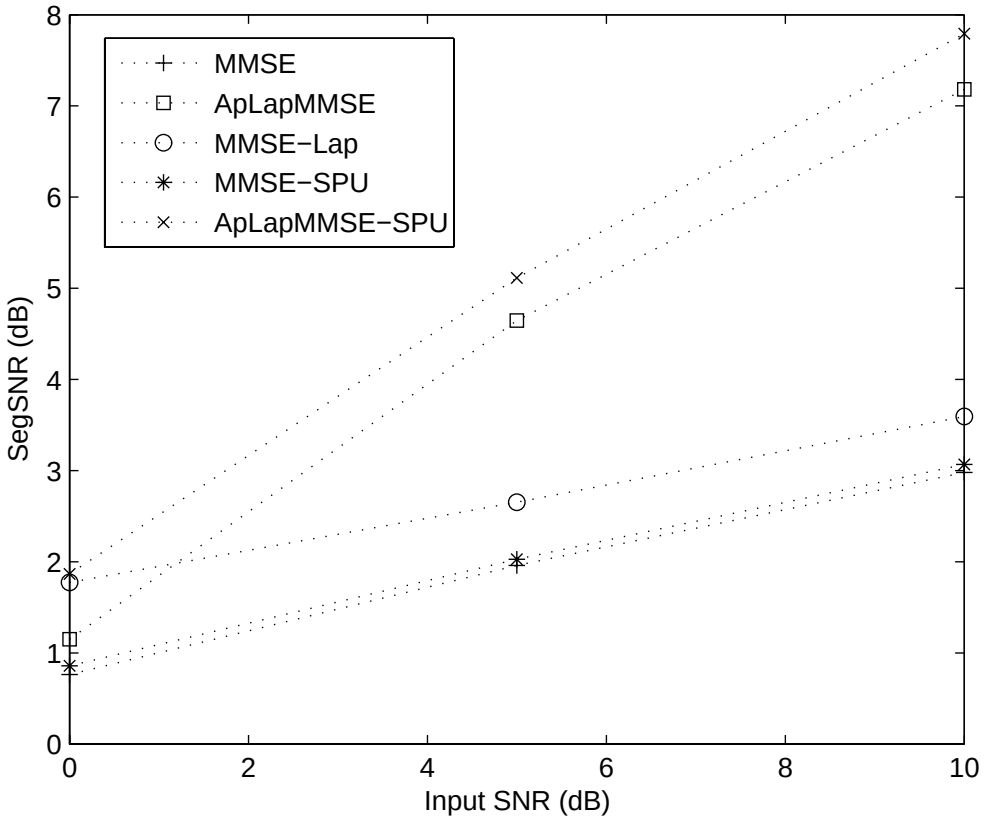

Figure 5.5. Comparative performance, in terms of segmental SNR, of the Gaussian-based MMSE and Laplacian-based MMSE estimators.

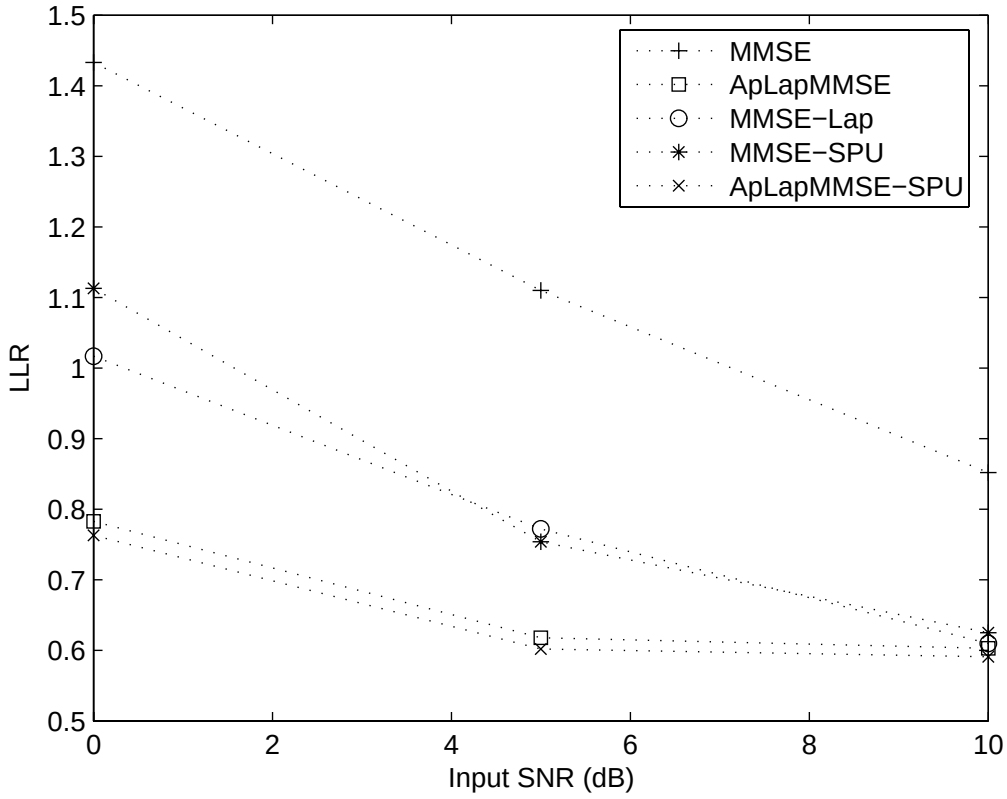

Figure 5.6. Comparative performance, in terms of log-likelihood ratio, of the Gaussian-based MMSE and Laplacian-based MMSE estimators.

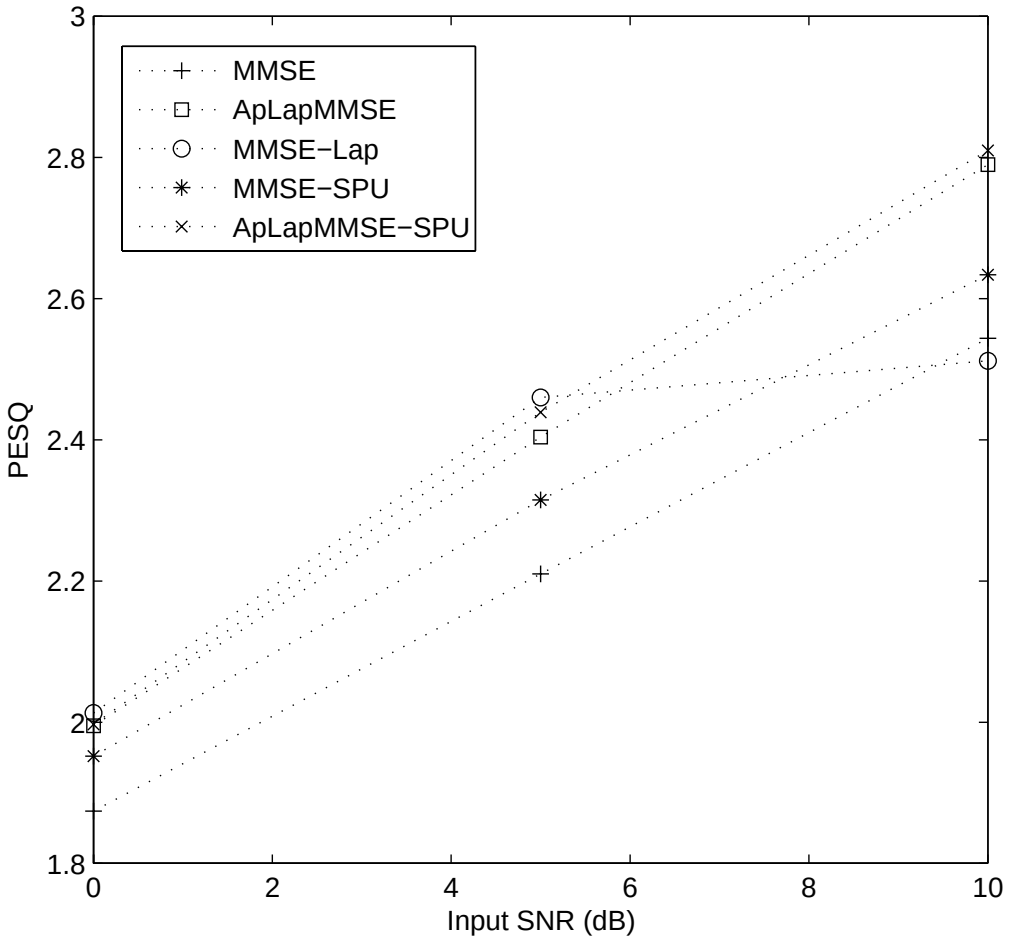

Figure 5.7. Comparative performance, in terms of PESQ, of the Gaussian-based MMSE and Laplacian-based MMSE estimators.

Estimator	Speech-Shaped			F-16 Fighter		
	0dB	5dB	10dB	0dB	5dB	10dB
MMSE	1.874	2.21	2.544	2.056	2.345	2.686
LapMMSE	1.882	2.448	2.783	2.19	2.587	2.914
ApLapMMSE	1.995	2.404	2.79	2.11	2.543	2.917
MMSE-Lap [35]	2.013	2.46	2.512	2.154	2.361	2.510
MMSE-SPU	1.952	2.315	2.634	2.03	2.346	2.70
LapMMSE-SPU	2.095	2.47	2.82	2.151	2.561	2.948
ApLapMMSE-SPU	1.997	2.439	2.81	2.146	2.563	2.943

Table 5.3. Comparative performance, in terms of PESQ, of the Gaussian-based MMSE and Laplacian-based MMSE estimators.

CHAPTER 6
SUMMARY AND CONCLUSIONS

An MMSE estimator was derived for the speech magnitude spectrum based on a Laplacian model for the speech DFT coefficients and a Gaussian model for the noise DFT coefficients. An estimator was also derived under speech presence uncertainty and a Laplacian model assumption. Results, in terms of objective measures, indicated that the proposed Laplacian MMSE estimators yielded better performance than the traditional MMSE estimator, which was based on a Gaussian model [1]. Overall, the present study demonstrated that the assumed distribution of the DFT coefficients could have a significant effect on speech quality.

Our main contributions are as follows:

1. The pdf of $Z = \sqrt{X^2 + Y^2}$, where X and Y follow a Laplacian pdf, was derived in closed form. The pdf was used in the derivation of the MMSE estimator based on Laplacian modeling of the speech DFT coefficients.

2. Much effort has been made to derive an MMSE estimator of the magnitude spectrum based on super-Gaussian priors. To our knowledge, no MMSE estimators of the magnitude spectrum based on Lalacian modeling has been derived. In this dissertation, we derive the optimal magnitude estimator of the speech spectrum

assuming a Laplacian distribution of the speech DFT coefficients and a Gaussian pdf for the noise DFT coefficients.

3. To further improve the quality of the enhanced speech, a Speech Presence Uncertainty (SPU) estimator based on Laplacian modeling is also proposed and derived. After the magnitude is estimated by the method proposed above, the SPU estimator refines the estimates of the magnitudes by scaling them with the SPU probability.

Future work can be pursued in the following directions:

1. Simplification of the pdf of spectral magnitude (Eq. (3.6)). The proposed MMSE magnitude estimator involves high computational complexity and potential numerical instability and needs to be handled carefully. If a simpler function is used to approximate the magnitude pdf (Eq. (3.6)) the high computational complexity and numerical instability issues may be circumvented. Also, better numerical computation techniques such as those in [60] [37] can be used.

2. Non-stationary noise. The proposed algorithms did not directly address non-stationary noise, and this is a very important issue for single channel speech enhancement. Noise estimation algorithms [61] [62] [63] can be incorporated in the proposed algorithms.

3. In this thesis, only Laplacian modeling of the speech DFT components was considered. Gamma modeling of the speech DFT coefficients is also encouraged.

APPENDIX

APPENDIX A

DERIVATION OF PDF OF SPECTRAL AMPLITUDE WITH COMPLEX LAPLACIAN DISTRIBUTED DFT COEFFICIENTS

This appendix details the derivation of the pdf of the random variable $Z = \sqrt{X_R^2 + X_I^2}$, where X_R, X_I are the real and imaginary parts respectively of DFT coefficients, assumed to have a *Laplacian* probability density function of the form:

$$p_{X_r}(x_r) = \frac{1}{2\sigma} \exp\left(-\frac{|x_r|}{\sigma}\right) \tag{A.1}$$

where σ is the standard deviation. Let $Y_R = X_R^2$ and $Y_I = X_I^2$. Then, we know [45] that $p_{Y_1}(y_1) = \frac{1}{\sqrt{y_1}} p_{x_1}(\sqrt{y_1})$ and so:

$$p_{Y_1}(y_1) = \frac{1}{\sqrt{y_1}} \frac{1}{2\sigma} \exp\left(-\frac{\sqrt{y_1}}{\sigma}\right) \ , \qquad p_{Y_2}(y_2) = \frac{1}{\sqrt{y_2}} \frac{1}{2\sigma} \exp\left(-\frac{\sqrt{y_2}}{\sigma}\right) \tag{A.2}$$

If we let $Z = X_R^2 + X_I^2 = Y_R + Y_I \quad (Y_R > 0, Y_I > 0)$, we know $p_Z(z) = p_{Y_R}(y_R) * p_{Y_I}(y_I)$, where $*$ indicates convolution. Thus,

$$\begin{aligned} p_Z(z) &= \int_{-\infty}^{\infty} p_{Y_R}(z - y_I) p_{Y_I}(y_I) dy_I \\ &= \int_{0+}^{z} \frac{1}{2\sigma\sqrt{z - y_R}} \exp(-\frac{\sqrt{z - y_I}}{\sigma}) \frac{1}{2\sigma\sqrt{y_I}} \exp(-\frac{\sqrt{y_I}}{\sigma}) dy_I \\ &= \frac{1}{4\sigma^2} \int_{0+}^{z} \frac{1}{\sqrt{y_I(z - y_I)}} \exp(-\frac{\sqrt{z - y_I} + \sqrt{y_I}}{\sigma}) dy_I \end{aligned} \tag{A.3}$$

Let $y_I = z \sin^2 t$

$$
\begin{aligned}
p_Z(z) &= \int_{0+}^{\pi/2} \frac{2z \sin t \cos t}{\sqrt{z^2 \sin^2 t (1 - \sin^2 t)}} \exp(-\frac{\sqrt{z}(\cos t + \sin t)}{\sigma}) dt \qquad \text{(A.4)} \\
&= \frac{1}{2\sigma^2} \int_0^{\pi/2} \exp(-\frac{\sqrt{2z} \cos(t - \pi/4)}{\sigma}) dt
\end{aligned}
$$

If we let $X = \sqrt{Z}$, we have $p_X(x) = 2x p_Z(x^2) u(x)$, where $u(x)$ is the step function. Substituting $\theta = t - \pi/4$ in the above equation, we get:

$$
\begin{aligned}
p_X(x) &= \frac{x}{\sigma^2} \int_0^{\pi/2} \exp(-\frac{\sqrt{2}x \cos(t - \pi/4)}{\sigma}) dt \qquad \text{(A.5)} \\
&= \frac{x}{\sigma^2} \int_{-\pi/4}^{\pi/4} \exp(-\frac{\sqrt{2}x \cos\theta}{\sigma}) d\theta \\
&= \frac{2x}{\sigma^2} \int_0^{\pi/4} \exp(-\frac{\sqrt{2}x \cos\theta}{\sigma}) d\theta \\
&= \frac{2x}{\sigma^2} \int_0^{\pi/4} \exp(Ax \cos\theta) d\theta \qquad\qquad x \geq 0
\end{aligned}
$$

where $A = -\sqrt{2}/\sigma$.

We tried to solve the above integral by expressing the exponential as an infinite series and then integrating term by term. We made use of the following generating function:

$$
\begin{aligned}
\exp\left[\frac{x}{2}(t + \frac{1}{t})\right] &= e^{xt/2} e^{x/2t} \qquad \text{(A.6)} \\
&= \left(\sum_{m=0}^{\infty} \frac{1}{m!} \frac{x^m t^m}{2^m}\right) \left(\sum_{n=0}^{\infty} \frac{1}{n!} \frac{x^n t^{-n}}{2^n}\right)
\end{aligned}
$$

For the above equation, we need to focus on the terms containing t^k $(k \geq 0)$. When they are multiplied together, a term containing t^n is obtained if the terms containing t^{-n} are multiplied by the terms in the first series containing t^{n+j}, that is, the terms for

which $m = k + n$. Therefore, taking into account all possible values of j, we find that the coefficient of t^n in the product of the two series is:

$$\sum_{n=0}^{\infty} \left[\frac{1}{(k+n)!}\frac{x^{k+n}}{2^{k+n}}\right]\left[\frac{1}{n!}\frac{x^n}{2^n}\right] = \sum_{n=0}^{\infty} \frac{x^{k+2n}}{2^{k+2n} n! \Gamma(k+n+1)} = I_k(x) \tag{A.7}$$

where $I_k(x)$ is the modified Bessel function of kth order. Similarly, for the terms containing t^{-k}, we get

$$\sum_{m=0}^{\infty} \left[\frac{1}{m!}\frac{x^m}{2^m}\right]\left[\frac{1}{(k+m)!}\frac{x^{k+m}}{2^{k+m}}\right] = \sum_{n=0}^{\infty} \frac{x^{k+2n}}{2^{k+2n} n! \Gamma(k+n+1)} = I_k(x) \tag{A.8}$$

Hence, summing all values of k, we get:

$$\exp\left[\frac{x}{2}(t + \frac{1}{t})\right] = I_0(x) + \sum_{k=1}^{\infty} I_k(x)[t^k + t^{-k}] \tag{A.9}$$

Let $t = e^{-i\theta}$, so that the left side of the above equation becomes:

$$\exp\left[\frac{x}{2}(t + \frac{1}{t})\right] = \exp\left[\frac{x}{2}(e^{-i\theta} + e^{i\theta})\right] = \exp(x\cos\theta) \tag{A.10}$$

and the right side becomes:

$$t^k + t^{-k} = e^{-ik\theta} + e^{ik\theta} = 2\cos k\theta \tag{A.11}$$

$$\begin{aligned} & I_0(x) + \sum_{k=1}^{\infty} I_k(x)[t^k + t^{-k}] \\ = \;& I_0(x) + 2\sum_{k=1}^{\infty} I_k(x)\cos k\theta \end{aligned} \tag{A.12}$$

Hence,

$$\exp(x\cos\theta) = I_0(x) + 2\sum_{k=1}^{\infty} I_k(x)\cos k\theta \tag{A.13}$$

Let $x = Ax$, so that

$$\begin{aligned} p_X(x) &= \frac{2x}{\sigma^2}\int_0^{\pi/4} \exp(Ax\cos\theta)d\theta \\ &= \frac{2x}{\sigma^2}\int_0^{\pi/4}\left(I_0(Ax) + 2\sum_{k=1}^{\infty} I_k(Ax)\cos k\theta\right)d\theta \\ &= \frac{2x}{\sigma^2}\left(\frac{\pi}{4}I_0(Ax) + 2\sum_{k=1}^{\infty} I_k(Ax)\int_0^{\pi/4}\cos k\theta d\theta\right) \\ &= \frac{2x}{\sigma^2}\left(\frac{\pi}{4}I_0(Ax) + 2\sum_{k=1}^{\infty}\frac{1}{k}I_k(Ax)\sin\frac{\pi n}{4}\right) \end{aligned}$$

Substituting $A = -\frac{\sqrt{2}}{\sigma}$, we have

$$p_X(x) = \frac{2x}{\sigma^2}\left[\frac{\pi}{4}I_0\left(-\frac{\sqrt{2}}{\sigma}x\right) + 2\sum_{k=1}^{\infty}\frac{1}{k}I_k\left(-\frac{\sqrt{2}}{\sigma}x\right)\sin\frac{\pi k}{4}\right] \tag{A.14}$$

If we let $\sigma^2 = \lambda$ and $A = -\sqrt{2/\lambda}$, then $p_X(x)$ becomes

$$p_X(x) = \frac{\pi x}{2\lambda}I_0\left(-\frac{\sqrt{2}}{\sqrt{\lambda}}x\right) + \frac{4x}{\lambda}\sum_{n=1}^{\infty}\frac{1}{n}I_n\left(-\frac{\sqrt{2}}{\sqrt{\lambda}}x\right)\sin\frac{\pi n}{4},\ \ x \geqslant 0 \tag{A.15}$$

Eq. (A.15) is the closed form solution of pdf of the magnitude desired.

APPENDIX B

DERIVATION OF APPROXIMATE LAPLACIAN MMSE ESTIMATOR

The MMSE amplitude estimator $\hat{X}_k$ can be expressed as

$$\begin{aligned} \hat{X}_k &= E\{X_k|Y(\omega_k)\} \\ &= \frac{\int_0^\infty \int_0^{2\pi} x_k p(Y(\omega_k)|x_k,\theta_k) p(x_k,\theta_k) d\theta_k dx_k}{\int_0^\infty \int_0^{2\pi} p(Y(\omega_k)|x_k,\theta_k) p(x_k,\theta_k) d\theta_k dx_k} \end{aligned} \tag{B.1}$$

where

$$p(Y(\omega_k)|x_k,\theta_k) = \frac{1}{\pi\lambda_d(k)} \exp\left\{ -\frac{1}{\lambda_d(k)} \left| Y_k - X_k e^{j\theta_x(k)} \right|^2 \right\}$$

Since $p(x_k,\theta_k) \approx \frac{1}{2\pi} p_X(x_k)$, we substitute Eq. (A.15) into Eq. (B.1) and then have

$$\begin{aligned} \hat{X}_k &= \frac{\int_0^\infty \int_0^{2\pi} x_k^2 \exp\left[-\frac{Y_k^2 - 2x_k Re\{e^{-j\theta_x} Y(\omega_k)\} + X_k^2}{\lambda_d(k)} \right] \Phi(x_k) d\theta_x dx_k}{\int_0^\infty \int_0^{2\pi} x_k \exp\left[-\frac{Y_k^2 - 2x_k Re\{e^{-j\theta_x} Y(\omega_k)\} + X_k^2}{\lambda_d(k)} \right] \Phi(x_k) d\theta_x dx_k} \\ &= \frac{\int_0^\infty x_k^2 \exp\left(-\frac{X_k^2}{\lambda_d(k)} \right) \Phi(x_k) \int_0^{2\pi} \exp\left(\frac{2x_k Re\{e^{-j\theta_x} Y(\omega_k)\}}{\lambda_d(k)} \right) d\theta_x dx_k}{\int_0^\infty x_k \exp\left(-\frac{X_k^2}{\lambda_d(k)} \right) \Phi(x_k) \int_0^{2\pi} \exp\left(\frac{2x_k Re\{e^{-j\theta_x} Y(\omega_k)\}}{\lambda_d(k)} \right) d\theta_x dx_k} \\ &= \frac{\int_0^\infty x_k^2 \exp\left(-\frac{X_k^2}{\lambda_d(k)} \right) I_0\left(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}} x_k \right) I_0(2x_k Y_k/\lambda_d(k)) + \Sigma(x_k) dx_k}{\int_0^\infty x_k \exp\left(-\frac{X_k^2}{\lambda_d(k)} \right) I_0\left(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}} x_k \right) I_0(2x_k Y_k/\lambda_d(k)) + \Sigma(x_k) dx_k} \end{aligned} \tag{B.2}$$

where

$$\Phi(x_k) = \left[I_0(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}} x_k) + \frac{8}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} I_n(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}} x_k) \sin\frac{\pi n}{4} \right]$$

$$\Sigma(x_k) = \frac{8}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{\pi n}{4}\int_0^{\infty} x_k^2 \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) I_n\left(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}}x_k\right) I_0(2x_kY_k/\lambda_d(k))$$

According to Eq. 6.633.1 in [39],

$$\int_0^{\infty} x^{\lambda+1}e^{-ax^2}J_\mu(\hat{\beta}x)J_\nu(\hat{\gamma}x)dx = \frac{\hat{\beta}^\mu\hat{\gamma}^\nu\alpha^{-\frac{\mu+\nu+\lambda+2}{2}}}{2^{\mu+\nu+1}\Gamma(\nu+1)}$$
$$\times\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}\nu+\frac{1}{2}\mu+\frac{1}{2}\lambda+1)}{m!\Gamma(m+\mu+1)}\left(-\frac{\hat{\beta}^2}{4\alpha}\right)^m$$
$$\times F(-m,-\mu-m;\nu+1;\frac{\hat{\gamma}^2}{\hat{\beta}^2})$$
$$\left[Re(\alpha)>0, Re(\mu+\nu+\lambda)>-2,\ \hat{\beta}>0,\ \hat{\gamma}>0\right]$$

Let $\hat{\beta} = j\beta$ and $\hat{\gamma} = j\gamma$, then the above equation becomes:

$$\int_0^{\infty} x^{\lambda+1}e^{-ax^2}J_\mu(j\beta x)J_\nu(j\gamma x)dx = \frac{j^{\mu+\nu}\beta^\mu\gamma^\nu\alpha^{-\frac{\mu+\nu+\lambda+2}{2}}}{2^{\mu+\nu+1}\Gamma(\nu+1)}$$
$$\times\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}\nu+\frac{1}{2}\mu+\frac{1}{2}\lambda+1)}{m!\Gamma(m+\mu+1)}\left(\frac{\beta^2}{4\alpha}\right)^m$$
$$\times F(-m,-\mu-m;\nu+1;\frac{\gamma^2}{\beta^2})$$

Converting $J_n(x)$ to $I_n(x)$ using $j^nI_n(x) = J_n(jx)$, we get:

$$\int_0^{\infty} x^{\lambda+1}e^{-ax^2}I_\mu(\beta x)I_\nu(\gamma x)dx = \frac{\beta^\mu\gamma^\nu\alpha^{-\frac{\mu+\nu+\lambda+2}{2}}}{2^{\mu+\nu+1}\Gamma(\nu+1)} \tag{B.3}$$
$$\times\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}\nu+\frac{1}{2}\mu+\frac{1}{2}\lambda+1)}{m!\Gamma(m+\mu+1)}\left(\frac{\beta^2}{4\alpha}\right)^m$$
$$\times F\left(-m,-\mu-m;\nu+1;\frac{\gamma^2}{\beta^2}\right)$$

We can use the above integral to evaluate Eq. (B.2) with the following arguments: $\lambda = 1$ or 0; $a = \frac{1}{\lambda_d(k)}$; $\beta = -\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}}$; $\gamma = \frac{2Y_k}{\lambda_d(k)}$; $\mu = 0$; $\nu = n$ $(0, 1, 2, 3, ...\infty)$.

Let us denote the numerator terms of Eq. (B.2) as A_k and B_k, respectively, i.e.,

$$A_k = \int_0^{\infty} x_k^2 \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) I_0\left(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}}x_k\right) I_0(2x_kY_k/\lambda_d(k))dx_k$$

and

$$B_k = \frac{8}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{\pi n}{4}\int_0^{\infty} x_k^2 \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) I_n\left(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}}x_k\right) I_0(2x_kY_k/\lambda_d(k))dx$$

and the denominator terms of Eq. (B.2) as C_k and D_k:

$$C_k = \int_0^{\infty} x_k \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) I_0\left(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}}x_k\right) I_0(2x_kY_k/\lambda_d(k))dx_k$$

and

$$D_k = \frac{8}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{\pi n}{4}\int_0^{\infty} x_k \exp\left(-\frac{X_k^2}{\lambda_d(k)}\right) I_n\left(-\frac{\sqrt{2}}{\sqrt{\lambda_x(k)}}x_k\right) I_0(2x_kY_k/\lambda_d(k))dx$$

Applying Eq. (B.3), we can evaluate the integrals of $A_k, B_k, C_k,$ and D_k:

$$A_k = \frac{\lambda_d(k)^{\frac{3}{2}}}{2}\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{3}{2})}{m!\Gamma(m+1)}\left(\frac{\lambda_d(k)}{2\lambda_x(k)}\right)^m F\left(-m,-m;1;2\frac{\lambda_x(k)}{\lambda_d^2(k)}Y_k^2\right) \tag{B.4a}$$

$$\begin{aligned} B_k &= \frac{8}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{\pi n}{4}\frac{\left(\frac{2Y_k}{\lambda_d(k)}\right)^n\left(\frac{1}{\lambda_d(k)}\right)^{-\frac{n+3}{2}}}{2^{n+1}\Gamma(n+1)}\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}n+\frac{3}{2})}{m!\Gamma(m+1)}\left(-\frac{\lambda_d(k)}{2\lambda_x(k)}\right)^m \\ &\cdot F\left(-m,-m;n+1;2\frac{\lambda_x(k)}{\lambda_d^2(k)}Y_k^2\right) \end{aligned} \tag{B.4b}$$

$$C_k = \frac{\lambda_d(k)}{2}\sum_{m=0}^{\infty}\frac{1}{m!}\left(\frac{\lambda_d(k)}{2\lambda_x(k)}\right)^m F\left(-m,-m;1;2\frac{\lambda_x(k)}{\lambda_d^2(k)}Y_k^2\right) \tag{B.4c}$$

$$\begin{aligned} D_k &= \frac{8}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{\pi n}{4}\frac{\left(\frac{2Y_k}{\lambda_d(k)}\right)^n\left(\frac{1}{\lambda_d(k)}\right)^{-\frac{n}{2}-1}}{2^{n+1}\Gamma(n+1)}\sum_{m=0}^{\infty}\frac{\Gamma(m+\frac{1}{2}n+1)}{m!\Gamma(m+1)}\left(\frac{\lambda_d(k)}{2\lambda_x(k)}\right)^m \\ &\cdot F\left(-m,-m;n+1;2\frac{\lambda_x(k)}{\lambda_d^2(k)}Y_k^2\right) \end{aligned} \tag{B.4d}$$

where $\Gamma(.)$ is the gamma function and $F(a,b,c;x)$ is the Gaussian hypergeometric function (Eq. (9.100) in [39]), which can be calculated using [60]. The above terms are

given as a function of the signal and noise variances, but can also be expressed in terms of the *a priori* *SNR* ξ_k ($\xi_k = \lambda_x(k)/\lambda_d(k)$) and *a posteriori* SNR γ_k ($\gamma_k = Y_k^2/\lambda_d(k)$) using the following relationships:

$$\frac{1}{\lambda_d(k)} = \frac{\gamma_k}{Y_k^2} \tag{B.5}$$

$$\frac{1}{\lambda_x(k)} = \frac{\gamma_k}{\xi_k\, Y_k^2} \tag{B.6}$$

Thus, we derive the MMSE estimator in closed form:

$$\hat{X}_k = \frac{A_k + B_k}{C_k + D_k} \tag{B.7a}$$

where

$$A_k = \frac{\left(\frac{Y_k^2}{\gamma_k}\right)^{\frac{3}{2}}}{2} \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{3}{2})}{m!\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m \cdot F\left(-m,-m;1;2\xi_k^2 Y_k^2\right) \tag{B.7b}$$

$$\begin{aligned} B_k &= \frac{8}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin\frac{\pi n}{4} \frac{\left(\frac{2\gamma_k}{Y_k}\right)^n \left(\frac{\gamma_k}{Y_k^2}\right)^{-\frac{n+3}{2}}}{2^{n+1}\Gamma(n+1)} \cdot \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}n+\frac{3}{2})}{m!\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m \\ &\cdot F\left(-m,-m;n+1;2\xi_k^2 Y_k^2\right) \end{aligned} \tag{B.7c}$$

$$C_k = \frac{Y_k^2}{2\gamma_k} \sum_{m=0}^{\infty} \frac{1}{m!} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m \cdot F\left(-m,-m;1;2\xi_k^2 Y_k^2\right) \tag{B.7d}$$

$$\begin{aligned} D_k &= \frac{8}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin\frac{\pi n}{4} \frac{\left(\frac{2\gamma_k}{Y_k}\right)^n \left(\frac{\gamma_k}{Y_k^2}\right)^{-\frac{n}{2}-1}}{2^{n+1}\Gamma(n+1)} \cdot \sum_{m=0}^{\infty} \frac{\Gamma(m+\frac{1}{2}n+1)}{m!\Gamma(m+1)} \left(\frac{\gamma_k}{2\xi_k^2 Y_k^2}\right)^m \\ &\cdot F\left(-m,-m;n+1;2\xi_k^2 Y_k^2\right) \end{aligned} \tag{B.7e}$$

Equation (B.7a) gives the approximate Laplacian MMSE estimator of the spectral magnitudes. We will be referring to this estimator as the ApLapMMSE estimator.

APPENDIX C

DERIVATION OF THE CONDITIONAL DENSITY $P(Y(\omega_K)|H_1^K)$

In this appendix, we compute $p(Y(\omega_k)|H_1^k)$, where $Y(\omega_k) = X(\omega_k) + D(\omega_k)$, assuming that the complex DFT coefficients of $X(\omega_k)$ and $D(\omega_k)$ follow a Laplacian and Gaussian distribution, respectively. Assuming independence between real and imaginary components, we have:

$$p(Y(\omega_k)|H_1^k) = p(z_r, z_i) = p_{Z_r(k)}(z_r)p_{Z_i(k)}(z_i) \tag{C.8}$$

where $Z_r(k) = Re\{Y(\omega_k)\}$, and $Z_i(k) = Im\{Y(\omega_k)\}$.

Let $Z_r = X_r + D_r$ and $Z_i = X_i + D_i$, then $Y = Z_r + jZ_i$. The pdf of Z_r can be computed by the convolution of the Laplacian and Gaussian densities, and is given by:

$$\begin{aligned} p_{Z_r}(z_r) &= \int_{-\infty}^{\infty} p_{X_r}(z_r - d_r)p_{D_r}(d_r)dd_r \\ &= \int_{-\infty}^{z_r} \frac{1}{2\sigma_x\sigma_d\sqrt{\pi}} \exp\left(-\frac{\sigma_x d_r^2 - \sigma_d^2 d_r + \sigma_d^2 z_r}{\sigma_x\sigma_d^2}\right) dd_r + \\ &\quad \int_{z_r}^{\infty} \frac{1}{2\sigma_x\sigma_d\sqrt{\pi}} \exp\left(-\frac{\sigma_x d_r^2 + \sigma_d^2 d_r - \sigma_d^2 z_r}{\sigma_x\sigma_d^2}\right) dd_r \end{aligned} \tag{C.9}$$

After using Eq. 2.338.1 in [39], we get:

$$\begin{aligned} p_{Z_r}(z_r) &= \frac{\exp(\frac{\sigma_d^2}{\sigma_x^2})}{2\sqrt{2}\sigma_x}[\exp(-\frac{z_r}{\sigma_x}) + \exp(\frac{z_r}{\sigma_x}) + \exp(-\frac{z_r}{\sigma_x})erf(\frac{z_r}{\sigma_x} - \frac{\sigma_d}{\sigma_x}) \\ &\quad - \exp(\frac{z_r}{\sigma_x})erf(\frac{z_r}{\sigma_x} + \frac{\sigma_d}{\sigma_x})] \end{aligned} \tag{C.10}$$

where $erf(\cdot)$ is the error function [44]. Finally, after expressing the above equation in terms of ξ_k and γ_k using the *a priori SNR* ξ_k ($\xi_k = \lambda_x(k)/\lambda_d(k)$) and *a posteriori* SNR γ_k ($\gamma_k = Y_k^2/\lambda_d(k)$), we get:

$$\begin{aligned} p_{Z_r(k)}(z_r) &= \frac{\sqrt{\gamma_k}\exp(\frac{1}{2\xi_k})}{2\sqrt{2\xi_k}Y_k}\left[\exp(-\frac{\sqrt{\gamma_k}z_r}{Y_k\sqrt{\xi_k}}) + \exp(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k}) + \right. \\ &\left. \exp(-\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k} - \frac{1}{\sqrt{\xi_k}}) - \exp(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_r}{\sqrt{\xi_k}Y_k} + \frac{1}{\sqrt{\xi_k}})\right] \end{aligned} \tag{C.11}$$

Similarly, the probability density for the imaginary part, i.e., $p_{Zi}(z_i)$, can be derived as follows:

$$\begin{aligned} p_{z_i}(z_i) &= \int_{-\infty}^{\infty} p_{X_r}(z_i - d_r)p_{D_r}(d_r)dd_r \\ &= \int_{-\infty}^{z_i} \frac{1}{2\sigma_x\sigma_d\sqrt{\pi}}\exp\left(-\frac{\sigma_x d_r^2 - \sigma_d^2 d_r + \sigma_d^2 z_i}{\sigma_x\sigma_d^2}\right)dd_r + \\ &\quad \int_{z_i}^{\infty} \frac{1}{2\sigma_x\sigma_d\sqrt{\pi}}\exp\left(-\frac{\sigma_x d_r^2 + \sigma_d^2 d_r - \sigma_d^2 z_i}{\sigma_x\sigma_d^2}\right)dd_r \end{aligned} \tag{C.12}$$

$$\begin{aligned} p_{z_i}(z_i) &= \frac{\exp(\frac{\sigma_d^2}{\sigma_x^2})}{2\sqrt{2}\sigma_x}[\exp(-\frac{z_i}{\sigma_x}) + \exp(\frac{z_i}{\sigma_x}) + \exp(-\frac{z_i}{\sigma_x})erf(\frac{z_i}{\sigma_x} - \frac{\sigma_d}{\sigma_x}) \\ &\quad - \exp(\frac{z_i}{\sigma_x})erf(\frac{z_i}{\sigma_x} + \frac{\sigma_d}{\sigma_x})] \end{aligned} \tag{C.13}$$

Expressing the above equation in terms of ξ_k and γ_k, we get the following expression for the conditional density $p(Y(\omega_k)|H_1^k)$ at frequency bin ω_k:

$$\begin{aligned} p_{z_i(k)}(z_i) &= \frac{\sqrt{\gamma_k}\exp(\frac{1}{2\xi_k})}{2\sqrt{2\xi_k}Y_k}\left[\exp(-\frac{\sqrt{\gamma_k}z_i}{Y_k\sqrt{\xi_k}}) + \exp(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k}) + \right. \\ &\left. \exp(-\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k} - \frac{1}{\sqrt{\xi_k}}) - \exp(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k})erf(\frac{\sqrt{\gamma_k}z_i}{\sqrt{\xi_k}Y_k} + \frac{1}{\sqrt{\xi_k}})\right] \end{aligned} \tag{C.14}$$

Substituting Eq. (C.11) and Eq. (C.14) to Eq. (4.7), we get $p(Y(\omega_k)|H_1^k)$.

APPENDIX D

NUMERICAL INTEGRATION TECHNIQUES BASED ON THE FINITE ELEMENT METHOD

Suppose we want to integrate a function $f(x)$ on $[a, b]$, which can be expressed as

$$f(x) = \sum_{m=0}^{M} f(x_m) V_m(x) \tag{D.1}$$

where $V_m(x)$ is defined as follows:

1.
$$V_0(x) = \begin{cases} \frac{x_0 - x}{d} \text{ for } x_0 < x < d + x_0 \\ 0 \qquad\qquad \text{Otherwise} \end{cases}$$

2.
$$V_m(x) = \begin{cases} \frac{x - x_{m-1}}{d} \text{ for } x_{m-1} < x \leqslant x_m \\ \frac{x - x_{m+1}}{-d} \text{ for } x_m \leqslant x < x_{m+1} \\ 0 \qquad\qquad \text{Otherwise} \end{cases}$$

3.
$$V_M(x) = \begin{cases} \frac{x - x_{M-1}}{d} \text{ for } x_{M-1} < x \leqslant x_M \\ 0 \qquad\qquad \text{Otherwise} \end{cases}$$

where $d = \frac{b-a}{M}$.

Integrating $f(x)$ we get:

$$\int_a^b f(x)dx = \sum_{m=0}^{M} f(x_m) \int_a^b V_m(x)dx = \frac{d}{2} f(x_0) + \frac{d}{2} f(x_M) + d \sum_{m=1}^{M-1} f(x_m) \tag{D.2}$$

For double integrals of the type $\int_C^D g(x)dx \int_a^b f(x,\theta)d\theta$, we can define:

$$\begin{aligned} F(x) &= \int_a^b f(x,\theta)d\theta = \int_a^b \sum_{m=0}^{M} f(x,\theta_m)V_m(\theta)d\theta \\ &= \frac{d}{2}\left[f(x,a)+f(x,b)\right] + d\sum_{m=0}^{M} f(x,\theta_m) \end{aligned} \tag{D.3}$$

where $\theta_m = a + md$ and $d = \frac{b-a}{M}$.

Next, let us choose an appropriate computation resolution M (trade-off between accuracy and computation load): $\Delta = \frac{D-C}{M}$, $x_m = C + \triangle m$. Then, we have

$$g(x) = \sum_{m=0}^{M} g(x_m)V_m(x)$$

$$F(x) = \sum_{m=0}^{M} F(x_m)V_m(x)$$

$$\int_C^D g(x)F(x)dx = \sum_{m=0}^{M}\sum_{l=0}^{M} g(x_m)F(x_l)\int_C^D V_l(x)V_m(x)dx \tag{D.4}$$

Note that $\int_C^D V_l(x)V_m(x)dx = 0$ if $l \neq m$ or $m-l \neq 1$, and the above integral is expressed as:

$$\begin{aligned}
\int_C^D g(x)F(x)dx &= g(x_0)F(x_0)\int_C^D v_0(x)v_0(x)dx + g(x_0)F(x_1)\int_C^D v_0(x)v_1(x)dx \qquad \text{(D.5)}\\
&\quad + \sum_{l=1}^{M-1}\left[g(x_l)F(x_l)\int_C^D v_l(x)v_l(x)dx + g(x_l)F(x_{l-1})\int_C^D v_l(x)v_{l-1}(x)dx \right.\\
&\quad \left. + g(x_{l+1})F(x_l)\int_C^D v_{l+1}(x)v_l(x)dx \right] + g(x_M)F(x_M)\int_C^D v_M^2(x)dx\\
&\quad + g(x_{M-1})F(x_M)\int_C^D v_M(x)v_{M-1}(x)dx\\
&= g(x_0)F(x_0)\int_C^D v_0(x)v_0(x)dx + g(x_0)F(x_1)\int_C^D v_0(x)v_1(x)dx\\
&\quad + \sum_{l=1}^{M-1}\left[g(x_l)F(x_l)\int_C^D v_l(x)v_l(x)dx + g(x_{l+1})F(x_l)\int_C^D v_{l+1}(x)v_l(x)dx \right]\\
&\quad + g(x_M)F(x_M)\int_C^D v_M^2(x)dx
\end{aligned}$$

The above three integrals evaluate to:

$$\int_C^D v_0(x)v_0(x)dx = \frac{\Delta}{3}$$

$$\int_C^D v_m(x)v_m(x)dx = \frac{2\Delta}{3}$$

$$\int_C^D v_m(x)v_{m+1}(x)dx = \frac{\Delta}{6}$$

Substituting the above three results into Eq. (D.5), we obtain the desired numerical integration formula:

$$\begin{aligned} & \int_C^D g(x)F(x)dx \qquad \text{(D.6)} \\ = \; & \frac{\Delta}{3}g(x_0)F(x_0) + \frac{\Delta}{6}g(x_0)F(x_1) + \frac{\Delta}{3}g(x_M)F(x_M) + \frac{\Delta}{6}g(x_{M-1})F(x_M) \\ & + \sum_{l=1}^{M-1}\left[\frac{2\Delta}{3}g(x_l)F(x_l) + \frac{2\Delta}{3}g(x_{l+1})F(x_l) + \frac{\Delta}{6}g(x_l)F(x_{l-1})\right] \end{aligned}$$

APPENDIX E
OBJECTIVE PERFORMANCE MEASURES

In this appendix, we describe the objective measures [52] used in this dissertation.

E.1 SNR

The SNR, or global SNR, is the most often used measure because it is simple to compute. Let $y(n)$, $x(n)$, and $d(n)$ denote the noisy speech signal, clean speech signal and noise signal, respectively, and $\hat{x}(n)$ the corresponding enhanced signal. The error signal $\varepsilon(n)$ can be written as

$$\varepsilon(n) = x(n) - \hat{x}(n) \tag{E.1}$$

The error signal energy can be computed as:

$$E_\varepsilon = \sum_n \varepsilon^2(n) = \sum_n [x(n) - \hat{x}(n)]^2 \tag{E.2}$$

and the signal energy as:

$$E_x = \sum_n x^2(n) \tag{E.3}$$

The resulting SNR measure (in dB) is obtained as

$$SNR = 10 \log_{10} \frac{E_x}{E_\varepsilon} = 10 \log_{10} \frac{\sum_n x^2(n)}{\sum_n [x(n) - \hat{x}(n)]^2} \tag{E.4}$$

E.2 Segmental SNR

The segmental SNR (SNR_{seg}) measure is a variation of SNR, and is formulated as follows

$$SNR_{seg} = \frac{1}{M}\sum_{j=0}^{M-1} 10\log_{10}\left[\frac{\sum_{n=m_j-N+1}^{m_j} x^2(n)}{\sum_{n=m_j-N+1}^{m_j}[x(n)-\hat{x}(n)]^2}\right] \tag{E.5}$$

where $m_0, m_1, \cdots, m_{M-1}$ are the end-times for the M frames, each of which is length N. For each frame, a SNR is computed and the final measure is obtained by averaging these measures over all segments of the waveform.

The main problem with the segmental SNR is that for some frames, the SNR value is unrealistically high or low, and that might severely bias the segmental SNR. To address this issue, two approaches can be taken, one is to set those unrealistically high or low SNR values to a high threshold (say 35dB) or a low threshold (say -10dB) respectively. The other approach is to totally abandon those unrealistically high or low SNR values. In this dissertation, we took the first approach.

E.3 LLR measure

The LLR (Log Likelihood Ratio) measure is a similarity measure based on Linear Prediction (LP) coefficients and is similar to the Itakura-Saito distance measure [64]. Specifically, for each frame m, we obtain the LP coefficients a_x of the clean signal and the LP coefficients a_d of the enhanced signal. The LLS measure is defined to be:

$$LLR = \log\left(\frac{a_d^T R_x a_d}{a_x^T R_x a_x}\right) \tag{E.6}$$

where a_x and a_d are the prediction coefficients of the clean and enhanced signals respectively, and R_x is the autocorrelation matrix of the clean signal. The mean LLR value was computed across all frames for each sentence.

APPENDIX F

MATLAB SOURCE CODE

F.1 Implementation of the MMSE Magnitude Estimator

The following code shows the major routine that implements the MMSE magnitude estimator in Eq. (3.9a). It calls three other functions listed in the following sections.

```
==================================================

    function [xhat]=MagEstimator(ksi,gammak,Yk,infi)

    %usage:  ksi: a priori SNR; gammak: a posterior SNR:

    %Yk: vector of DFT coefficients of noisy spectrum

    %infi: a value used to approximate infinity (here infi=40)

    Yk2=Yk.*Yk;

    L=length(ksi)/2+1;

    d=infi/200;

    x=[0:d:infi];

    for j=1:L

        % Calculate pdf of the modulos Z = sqrt(X^2 + Y^2)
```

```
        fn1=pdf(-1.414*gammak(j)/ksi(j)/Yk2(j),x);
        fn2=(x.^2).*exp((-gammak(j)/Yk2(j))*...
        ...x.^2).*besseli(0,(2*gammak(j)/Yk(j))*x);
        fn3=x.*exp((-gammak(j)/Yk2(j))*x.^2).*...
        ...besseli(0,(2*gammak(j)/Yk(j))*x);
        fn2(find(isnan(fn2)==1))=eps; %workaround for the case that inf*infsmall
        fn3(find(isnan(fn3)==1))=eps; %it's set to a very small value because
                                      % infsmall is of higher order over inf.
        A(j)=dblint(fn1,fn2,d); % Integration of multiplication of two functions
        C(j)=dblint(fn1,fn3,d);
        xhat(j,1)=A(j)/C(j);      % Eq. (3.9a)
    end
    xhat = [xhat; flipud(xhat(2:L-1))];
```

==========

F.2 Subroutine for Calculation of pdf of $X = \sqrt{X_R^2 + X_I^2}$

The following code computes the pdf of X where $X = \sqrt{X_R^2 + X_I^2}$ in Eq. (3.6) assuming X and Y follow a Laplacian distribution.

```
==========================================================

    function y=pdf(a,w)

    % y = p(ω)

    % a = −√2/σ where σ is the standard deviation of the signal.

    % w is a vector containing the range of values to evaluate the pdf.

    d=0.01;

    inv=[0:d:pi/4];

    N=length(inv);

    dw=w(2)-w(1);

    k=1;

    for w=min(w):dw:max(w)

      fun=exp(a*w*cos(inv));

      y(k)=8*w*d*(sum(fun)-(fun(1)+fun(N))/2);      %Eq. (3.6)

      k=k+1;

    end

==========================================================
```

F.3 Subroutine for Numerical Integration of $\int_a^b f(x)g(x)dx$

The following code implements the numerical integration of $\int_a^b f(x)g(x)dx$, where fn1 and fn2 represent $f(x)$ and $g(x)$, respectively. It uses Eq. (5.1).

```
==================================================

    function y=dblint(fn1,fn2,d);

    % Integration of product of two functions

    N=length(fn1);

    zero_zero=d*(fn1*fn2')/3*2-d*(fn1(1)*fn2(1)+fn1(N)*fn2(N))/3;

    zero_one=d*(fn1(1:N-1)*fn2(2:N)')/6;

    one_zero=d*(fn1(2:N)*fn2(1:N-1)')/6;      %Eq. (5.1)


    y=zero_zero+zero_one+one_zero;
==================================================
```

F.4 Subroutine for SPU Estimator

This code implements the SPU Estimator in Eq. (4.7) The *posterior* SNR is limited to a maximum of 20dB to avoid numerical instability.

```
==================================================

    function p=spchdector(Yk,gammak,ksi,Ndelta)
```

```
%usage:  ksi: a priori SNR; gammak: a posterior SNR:
%Yk: vector of DFT coefficients of noisy spectrum
%Ndelta: Noise variance
Ykr=real(Yk);
Yki=imag(Yk);
Ykm=abs(Yk);
SNRr=sqrt(gammak./ksi);
commr=Ykr.*SNRr./Ykm;
commi=Yki.*SNRr./Ykm;
sqrt2=sqrt(2);
qk=0.3;
fnr=SNRr.*exp(0.5./ksi).*(exp(-commr).*(1+erf(commr-1./sqrt(ksi)))+...
exp(commr).*(1-erf(commr+1./sqrt(ksi)))./(2*sqrt2*Ykm);  %Eq. (4.8)
fni=SNRr.*exp(0.5./ksi).*(exp(-commi).*(1+erf(commi-1./sqrt(ksi)))+...
exp(commi).*(1-erf(commi+1./sqrt(ksi)))./(2*sqrt2*Ykm); %Eq. (4.9)
% cap Posterior SNR below 20dB to avoid underflow in the next step
pSNR=min(100,(Ykm.^2)./Ndelta);
fns=exp(-pSNR)./(pi*Ndelta);       %Eq. (4.8)
```

```
A=(1/qk-1)*(fnr.*fni./fns);

p=A./(1+A);

p(find(isnan(p)==1))=1;
```

==========

REFERENCES

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

[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] F. Toledo, P. C. Loizou, and A. Lobo, "Subspace and envelope subtraction algorithms for noise reduction in cochlear implants," in *IEEE Annual International Conference of Engineering Medicine and Biology Society*, pp. 213–216, 2003.

[4] M. Rosenblatt, *Time Series Analysis*. Wiley, 1963.

[5] S. Gustafsson, P. Jax, and P. Vary, "A novel psychoacoustically motivated audio enhancement algorithm preserving background noise characteristics," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 397–400, 1998.

[6] F. Jabloun and B. Champagne, "A perceptual signal subspace approach for speech enhancement in colored noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 569–572, 2002.

[7] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech Audio Proc.*, vol. 7, pp. 126–137, Mar. 1999.

[8] P. M. Crozier, B. Cheetham, C. Holt, and E. Munday, "Speech enhancement employing spectral subtraction and linear predictive analysis," *Electronics Letters*, vol. 29, pp. 1094–1095, June 1993.

[9] L. Lin, W. H. Holmes, and E. Ambikairajah, "Adaptive noise estimation algorithm for speech enhancement," *Electronics Letters*, vol. 39, pp. 754–755, May 2003.

[10] L. Deng, J. Droppo, and A. Acero, "Recursive estimation of nonstationary noise using iterative stochastic approximation for robust speech recognition," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 11, pp. 568–580, Nov. 2003.

[11] V. Mathews, Y. Dae, and N. Ahmed, "A unified approach to nonparametric spectrum estimation algorithms," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 35, pp. 338–349, Mar. 1987.

[12] R. Martin, "Speech enhancement using mmse short time spectral estimation with gamma distributed speech priors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 253–256, 2002.

[13] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-33, pp. 443–445, 1985.

[14] R. Martin, "Spectral subtraction based on minimum statistics," in *Proc. Euro. Signal Processing Conf.(EUSIPCO)*, pp. 1182–1185, 1994.

[15] J. S. Lim and A. V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 26, pp. 197–210, 1978.

[16] R. J. McAulay and M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 28, pp. 137–145, Apr. 1980.

[17] M. S. Ahmed, "Comparison of noisy speech enhancement algorithms in terms of lpc perturbation," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 37, pp. 121–125, 1989.

[18] H. R. Abutalebi and H. Sheikhzadeh, "A hybrid subband adaptive system for speech enhancement in diffuse noise fields," *IEEE Signal Processing Letters*, vol. 17, pp. 270–273, 2004.

[19] A. T. Walden, D. B. Percival, and E. J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 3153–3165, 1998.

[20] B. Carnero and A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithms," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 47, pp. 1622–1635, 1999.

[21] P. Wolfe and S. Godsill, "Simple alternatives to the ephraim and malah suppression rule for speech enhancement," *Proc. IEEE Workshop on Statistical Signal Processing*, vol. 11th, pp. 496–499, Aug. 2001.

[22] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, pp. 2403–2418, Jan. 2001.

[23] R. Martin, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 504–512, 2002.

[24] D. R. Brillinger, *Time Series: Data Analysis and Theory.* San Francisco: Holden Day, Inc., 1981.

[25] J. Porter and S. Boll, "Optimal estimators for spectral restoration of noisy speech," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 18A.2.1–18A.2.4, 1984.

[26] C. Breithaupt and R. Martin, "Mmse estimation of magnitude-squared dft coefficients with supergaussian priors," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 848–851, 2003.

[27] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 208–211, 1979.

[28] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-27, pp. 113–120, 1979.

[29] J. M. Tribolet and R. E. Crochiere, "Frequency domain coding of speech," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-27, p. 522, Oct. 1979.

[30] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-25, p. 306, Aug. 1977.

[31] W. B. Davenport and W. L. Root, *An Introduction to the theory of random signals and noise.* New York: McGraw-Hill, 1958.

[32] H. Brehm and W. Stammler, "Description and seneration of spherically invariant speech-model signals," *Signal Processing*, vol. 12, pp. 119–141, 1987.

[33] R. Martin, "Speech enhancement based on minimum mean-squareestimation and supergaussian priors," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 13, pp. 845–856, Sept. 2005.

[34] S. Kullback, *Information theory and statistics.* Dover Publication, 1958.

[35] R. Martin and C. Breithaupt, "Speech enhancement in the dft domain using laplacian speech priors," in *International Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 87—90, Sept. 2003.

[36] T. Lotter and P. Vary, "Noise reduction by maximum a posteriori spectral amplitude estimation with supergaussian speech modeling," in *Proc. Int. Workshop*

Acoustic Echo and Noise Control, pp. 83–86, 2003.

[37] K. E. Mueller, "Computing the confluent hypergeometric function, m(a, b, x)," *Numer. Math.*, vol. 90, pp. 179–196, 2001.

[38] A. J. Accardi and R. V. Cox, "A modular approach to speech enhancement with an application to speech coding," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1999.

[39] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*. Academic Press, 2000.

[40] P. Vary, "Noise suppression by spectral magnitude estimation–mechanism and theoretical limits," *Signal Processing*, vol. 8, pp. 387–400, 1985.

[41] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, Sept. 1982.

[42] A. C. Cristán and A. T. Walden, "Multitaper power spectrum estimation and thresholding: Wavelet packets versus wavelets," *IEEE Trans. Signal Processing*, vol. 50, pp. 2976–2986, Dec. 2002.

[43] S. M. Kay, *Modern Spectral Estimation*. NJ: Prentice Hall, 1988.

[44] M. Loeve, *Probability neory*. Princeton, third ed., 1963.

[45] A. Papoulis and U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2001.

[46] D. Middleton and R. Esposito, "Simultanious optimum detection and estimation of signals in noise," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 434–444, May 1968.

[47] D. Malah, R. V. Cox, and A. J. Accardi, "Tracking speech-presence uncertainty to improve speech enhancement in non-stationary noise environment," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 1, pp. 789–792, 1999.

[48] I. Cohen, "Optimal speech enhancement under signal presence uncertainty using log-spectral amplitude estimator," *IEEE Signal Processing Lett.*, vol. 9, pp. 113–116, Apr. 2002.

[49] J. Yang, "Frequency domain noise suppression approaches in mobile telephone systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 363–366, 1993.

[50] Y. Kwon and H. Bang, *The Finite Element Method Using Matlab.* New York: CRC Press, 2000.

[51] D. B. Paul, "The spectral envelope estimation vocoder," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-29, pp. 786–794, 1981.

[52] J. R. Deller, J. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals.* NY: IEEE Press, 2000.

[53] M. R. Portnoff, "Time frequency representation of digital signals and systems based on short-time fourier analysis," *IEEE Trans. Acoustics, Speech and Signal*

Processing, vol. ASSP-28, pp. 55–69, 1980.

[54] R. E. Crochiere, "A weighted overlap-add method of short-time fourier analysis/synthesis," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-28, pp. 99–102, 1980.

[55] S. Quackenbush, T. Barnwell, and Clements, *Objective measures of speech quality.* Englewood Cliffs, NJ: Prentice-Hall, 1988.

[56] I.-T. P.862, *Perceptual evaluation of speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.* ITU-T Recommendation P.862, 2000.

[57] T. Goldstein and A. W. Rix, "Perceptual speech quality assessment in acoustic and binaural applications," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 1064–1067, 2004.

[58] J. Hansen and B. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *Int. Conf. Spoken Language Processing*, pp. 2819–2822, Dec. 1998. Sydney, Australia.

[59] P. D. Souza and P. Thomson, "Perceptual coding of digital audio," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 30, pp. 304–315, 1982.

[60] M. Nardin, W. Perger, and A. Bhalla, "Numerical evaluation of the confluent hypergeometric function for complex arguments of large magnitudes," *J. Comput. Appl. Math.*, vol. 39, pp. 193–200, 1992.

[61] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Proc.*, vol. 9, pp. 504–512, July 2001.

[62] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Lett.*, vol. 9, pp. 12–15, Jan. 2002.

[63] S. Rangachari, P. C. Loizou, and Y. Hu, "A noise estimation algorithm with rapid adaptation for highly nonstationary environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp. 305–308, 2004.

[64] P. Chu and D. Messerschmitt, "A frequency weighted itakura-saito spectral distance measure," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 30, pp. 545–560, Aug. 1982.

VITA

Bin Chen was born in Qingdao City, China in 1973, the son of Jirong Chen and Xinjie Yu. After graduation from Qingdao City No. 1 High School in 1992, he attended University of Qingdao. He graduated in 1996 with a Bachelor of Science degree in Electrical Engineering. Also in 1996, he attended Ocean University of China and graduated with a Master of Science degree in Electrical Engineering in 1999. In the fall of 2001, he began his PhD studies in Electrical Engineering at the University of Texas at Dallas, with a research assistantship from an NIH grant. He married Yan Bai in Jan of 2001.