DISTRIBUTED MULTICHANNEL PROCESSING FOR SIGNAL ENHANCEMENT

by

Marek B. Trawicki, B.S.E.E., M.S.E.E., M.S.

A Dissertation submitted to the Faculty of the
Graduate School, Marquette University,
in Partial Fulfillment of
the Requirements for
the Degree of
Doctor of Philosophy

Milwaukee, Wisconsin

May 2009

ABSTRACT
DISTRIBUTED MULTICHANNEL PROCESSING FOR SIGNAL ENHANCEMENT

Marek B. Trawicki, B.S.E.E., M.S.E.E., M.S.

Marquette University, 2009

The goal of this work is to generalize speech enhancement methods from single channel microphones, dual channel microphones, and microphone arrays to distributed microphones. The focus has been on developing and implementing robust and optimal time domain and frequency domain estimators for estimating the true source signal in this configuration and measuring the performance improvement with both objective (e.g., signal-to-noise ratios) and subjective (e.g., listening tests) metrics. Statistical estimation techniques (e.g., minimum mean-square error or MMSE) with Gaussian speech priors and Gaussian noise likelihoods have been used to derive solutions for five basic classes of estimators: 1) time domain; 2) spectral amplitude; 3) perceptually-motivated spectral amplitude; 4) spectral phase; and 5) complex real and imaginary spectral component. Experimental work using different true source signal attenuation factors (e.g., unity, linear, and logarithmic) demonstrates significant gains in segmental signal-to-noise ratio (SSNR) with an increase in the number of microphones. Of particular importance is the inclusion of the optimal MMSE spectral phase estimator to the spectral amplitude estimators. Overall, the statistical estimators show tremendous promise for distributed microphone speech enhancement of noisy acoustic signals with application to many consumer, industrial, and military products under severely noisy environments.

ACKNOWLEDGEMENTS

Marek B. Trawicki, B.S.E.E., M.S.E.E., M.S.

I wish to thank my adviser, Dr. Michael T. Johnson, for all of his guidance and support during my research, studies, and teaching experiences. I appreciate all of the generous time and patience that he has extended towards me. I am grateful that he was always available to discuss my new ideas, clarify any questions about theoretical concepts, review my mathematical equations, and troubleshoot code with me. I am thankful that he allowed me to learn from him with the hopes of becoming as quality and caring of a professor someday as him. I wish to thank him again for all of his dedication, inspiration, and expertise he has shared with me.

I wish to thank my committee members Dr. Richard J. Povinelli, Dr. James E. Richie, Dr. Edwin E. Yaz, and Dr. Fabien J. Josse for being so helpful and gracious with their comments and suggestions on my research work and serving on my committee.

I wish to thank the National Science Foundation and the U.S. Department of Education for all of their financial support.

I wish to thank all of my classmates and professors at Marquette University who always made themselves available for enriching and engaging conversations involving numerous homework assignments, projects, and research topics.

I wish to thank my colleagues at Marquette University in the Speech and Signal Processing laboratory, especially Pat, Kun, Li, Yao, Jidong, and Jianglin. I enjoyed our

many interesting and engaging discussions and social gatherings and learned so much from them. I look forward to hopefully further collaborations with them in the future.

I wish to thank all of my students, tutors, staff, and friends whom I have tutored and met over the last four years at OSES, especially Dawn, Karen, Pat, Heidi, and Jean. I had enjoyed so many wonderful and humorous conversations with all of them, and I was able to further develop myself as a person and learn a great deal. I will always cherish the amazing relationships that I have developed with everyone.

I wish to thank all of my family and friends, especially my brother, Richard. I am happy that he was always around to help me greatly balance my academic and social life. I enjoyed all of our exciting and fun adventures in the lab and around the town with him. I was better able to appreciate the successes of my work with the joy of his wonderful friendship and company.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

In this chapter, the problem of distributed microphone speech enhancement is introduced and related to prior work with single channel microphones, dual channel microphones, and microphone arrays. Several microphone configuration scenarios are compared in terms of effectiveness of current standard methods.

1.1. Problem Statement

Over the past several decades, there has been a great deal of research in the signal processing community on the development and implementation of speech enhancement algorithms. Whereas the current state-of-the-art methods work reasonably well for some applications, the performance of the algorithms quickly deteriorates under highly noisy conditions. In order to improve quality and intelligibility of speech enhancement systems, researchers have begun to investigate the use of multichannel (dual, array, and distributed) microphones to reduce noise in noisy speech signals and exploit all available acoustic and spatial information of the speech and noise sources [1]. Figure 1-1. Figure 1-2, Figure 1-3, and Figure 1-4 compare the various microphone configurations for clean speech s, noise n, attenuation factors c_i,time delays τ_i, and microphones M_i.

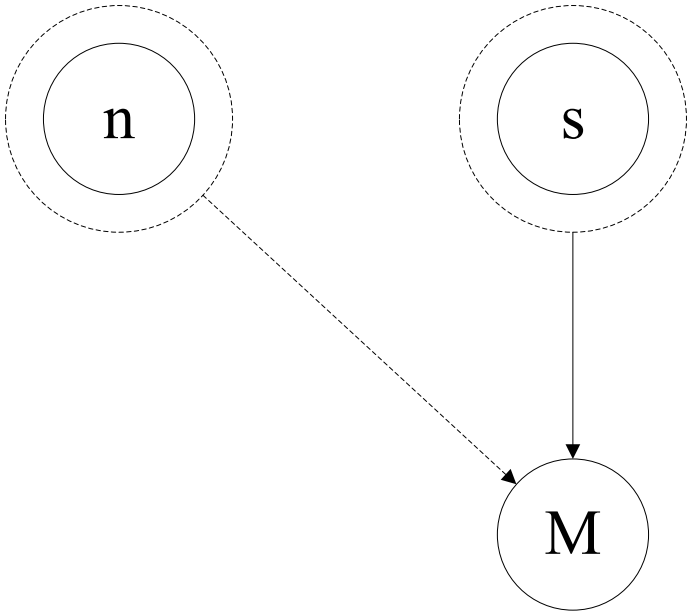

Figure 1-1 Single Channel Microphones

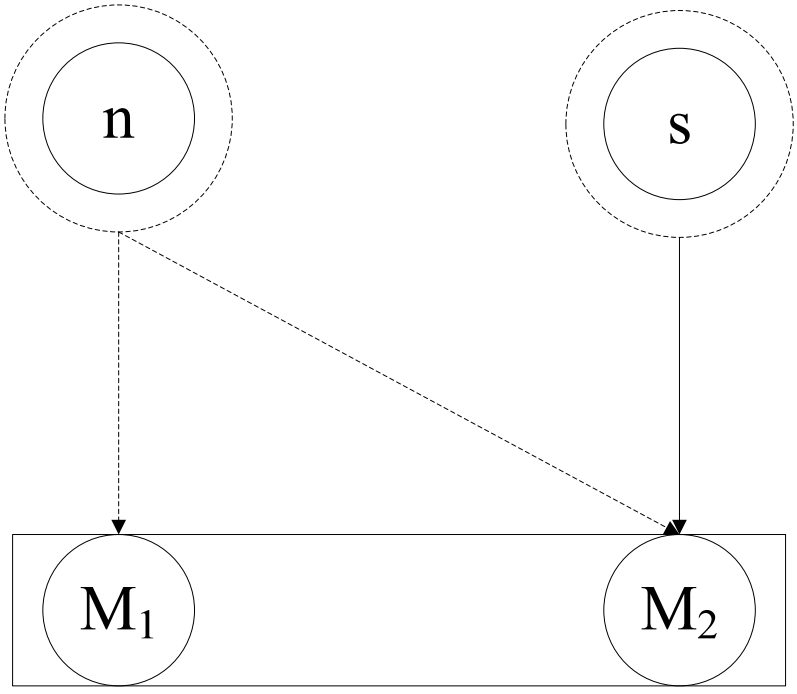

Figure 1-2 Dual Channel Microphones

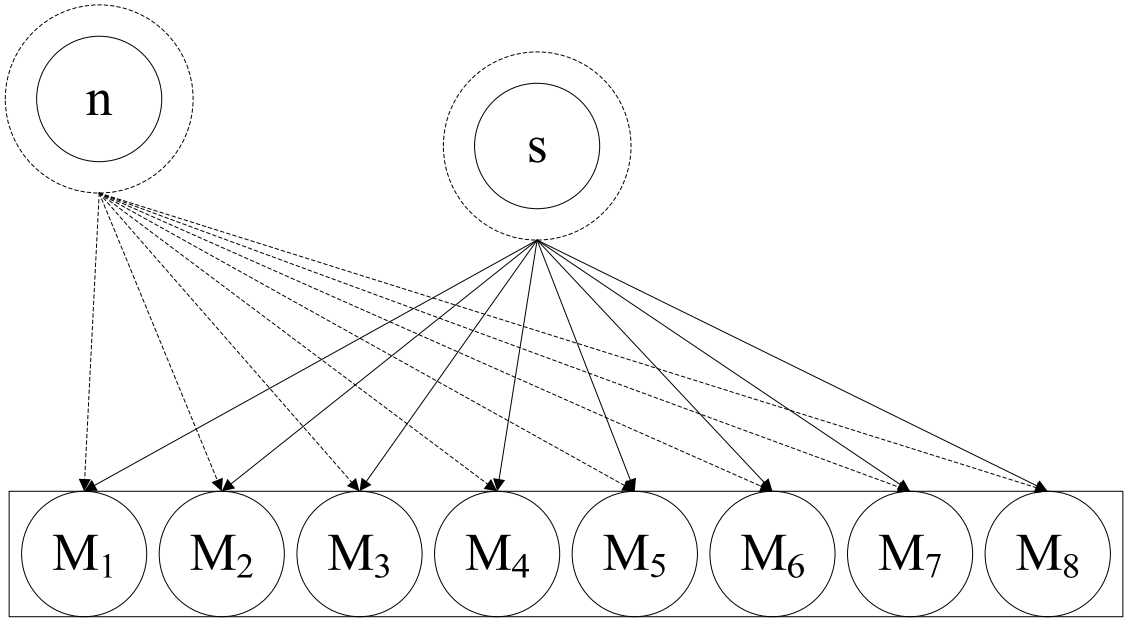

Figure 1-3 Microphone Array

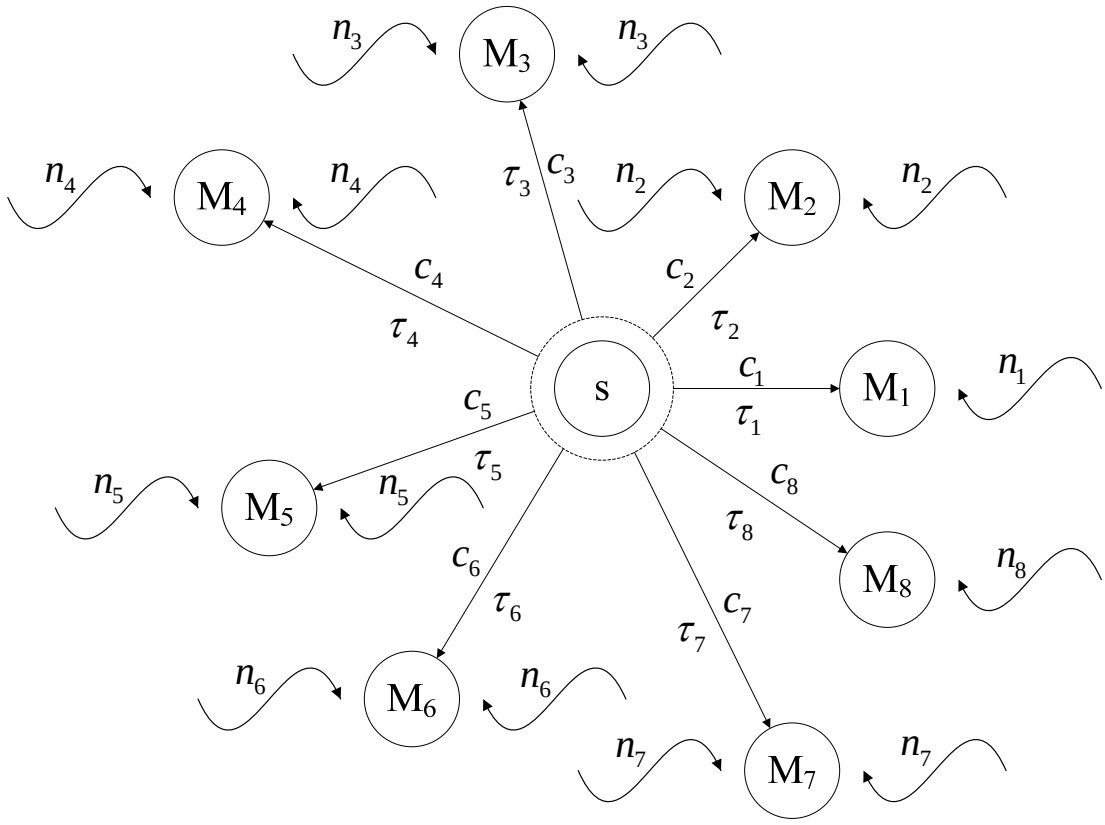

Figure 1-4 Distributed Microphones

While single channel microphones require the speakers to be relatively close to the microphone and dual channel microphones involve a reference noise microphone [2], microphone arrays [3] necessitate close-spacing of the microphones and *a priori* knowledge of the array geometry with the distances between individual array elements small enough to allow for spatial signal processing without aliasing and justify assumptions of noise correlation across the channels [2, 4-8]. Distributed microphones generalize single channel microphones, dual channel microphones, and microphone arrays. Speakers may be far away from the microphones, which are spread throughout a large area with unknown spacing and configurations and no longer satisfy the array assumptions. Table 1-1 provides a summary of the different microphone scenarios.

Configuration	Comments	Performance	Application
Single Channel Microphones	Common in practice Requires subjects close to the microphone Extensive research and existing algorithms	Speech enhancement degrades significantly in the presence of background noise	Hands-free and mobile communication
Dual Channel Microphones	Common in practice Requires reference microphone with only noise Significant research and existing algorithms	Speech enhancement improves over single channel microphone in presence of background noise	Hearing aids
Microphone Arrays	Common in practice Requires close microphone spacing and *a priori* knowledge of geometry Good deal of research and existing algorithms	Speech enhancement improves over single channel microphone in presence of background noise	Hearing aids
Distributed Microphones	Becoming more common in practice Allows arbitrary placement of microphones Not nearly as much research and as many existing algorithms	Speech enhancement improves over single channel microphone in presence of background noise	Speaker spotting, identification, and tracking systems

Table 1-1 Microphone Configurations

There has been relatively little work for distributed microphone speech enhancement compared to single channel microphones, dual channel microphones, and microphone arrays. Table 1-2 shows the common methods for performing speech enhancement for each of the microphone configurations.

<table>
<tr><th>Method</th><th>Single Channel Microphones</th><th>Dual Channel Microphones</th><th>Microphone Arrays</th><th>Distributed Microphones</th></tr>
<tr><td rowspan="6">Speech Enhancement</td><td>Spectral Subtraction [9]</td><td rowspan="6">Adaptive Noise Cancellation [2]</td><td rowspan="3">Fixed Beamforming [3]</td><td rowspan="3">Wiener Filter [10]</td></tr>
<tr><td>Wiener Filter [11]</td></tr>
<tr><td>Short-Time Spectral Amplitude Estimation [5]</td></tr>
<tr><td>Log-Spectral Amplitude Estimation [6]</td><td rowspan="3">Adaptive Beamforming [3]</td><td rowspan="3">Short-Time Spectral Amplitude Estimation [7]</td></tr>
<tr><td>Perceptually-Motivated Spectral Amplitude Estimation [12]</td></tr>
<tr><td>Complex Real and Imaginary Spectral Component Estimation [13]</td></tr>
</table>

Table 1-2 Traditional Methods for Speech Enhancement

In order to advance the current state-of-the-art speech enhancement methods for distributed microphones, it is important to generalize the existing work from single channel microphones, dual channel microphones, and microphone arrays.

1.2. Research Objectives

The ultimate goal of this research is to develop and implement a novel framework through statistical estimation [14, 15] for performing distributed microphone speech enhancement on noisy speech signals. The distributed microphone statistical estimators are categorized into the following classes:

1. Time Domain Estimation
2. Spectral Amplitude Estimation
3. Perceptually-Motivated Spectral Amplitude Estimation
4. Spectral Phase Estimation
5. Complex Real and Imaginary Spectral Component Estimation

For the spectral amplitude estimators, the key component to improvements in quality and intelligibility is due to the spectral phase estimator. Overall, the derived systems have the ability to estimate the true source signal with application to many consumer, industrial, and military products.

1.3. Dissertation Overview

The remainder of this dissertation is organized into the following sections: Background (CHAPTER 2), Theoretical Methods (CHAPTER 3), Experimental Work (CHAPTER 4), and Conclusion (CHAPTER 5).

CHAPTER 2 BACKGROUND

In this chapter, the fundamental concepts and standard methods are introduced for speech enhancement involving various microphone configuration scenarios: single channel microphones, dual channel microphones, microphone arrays, and distributed microphones. Noise estimation techniques, which are a central element of the speech enhancement estimators, are discussed along with an overview of the speech enhancement process. With each of the methods, the mathematics are highlighted in both the time domain and frequency domain along with appropriate performance evaluation metrics.

2.1. Overview

The goal of speech enhancement is to increase both the quality and intelligibility of the noisy speech signals. Figure 2-1 shows the basic process of performing speech enhancement on the single channel production model that consists of the clean speech signal s with uncorrelated additive noise d.

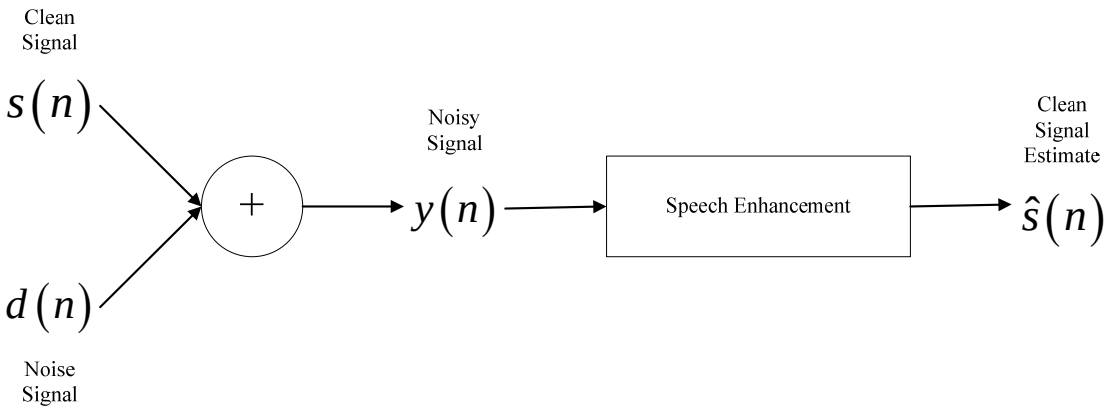

Figure 2-1 Speech Enhancement Applied to Single Channel Production Model

From Figure 2-1, the basic time domain single channel model is

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

or, in the discrete time domain

$$y(n) = s(n) + d(n), \tag{2.2}$$

where $s(n)$, $d(n)$, and $y(n)$ are the clean signal, noise signal, and noisy signal at time $t = nT$. Application of a speech enhancement algorithm to the noisy signal $y(n)$ creates an estimate of the clean signal $\hat{s}(n)$.

Performance can be measured by either objective or subjective metrics. Whereas the objective quality is measured through the output signal-to-noise ratio (SNR) and segmental signal-to-noise ratio (SSNR), the subjective quality is measured through listening tests such as the widely-used Mean Opinion Score [16]. The objective quality metrics of SNR and SSNR are defined as

$$SNR = 10\log_{10}\left(\frac{\sum_n s^2(n)}{\sum_n \left[s(n) - \hat{s}(n)\right]^2}\right) \tag{2.3}$$

and

$$SSNR = \frac{1}{M}\sum_M 10\log_{10}\left(\frac{\sum_n s^2(n)}{\sum \left[s(n) - \hat{s}(n)\right]^2}\right) \tag{2.4}$$

for frames $m \in M$. For subjective quality metrics, listening test subjects are asked to assign a score from 1-5 to each of the speech signals. Table 2-1 describes the rankings for MOS.

Rating	Speech Quality	Level of Distortion
5	Excellent	Imperceptible
4	Good	Just perceptible but not annoying
3	Fair	Perceptible and slightly annoying
2	Poor	Annoying but not objectionable
1	Unsatisfactory	Very annoying and objectionable

Table 2-1 MOS Five-Point Scale

In order to improve either the objective quality or subjective quality, the crucial factor in any speech enhancement system is the ability to accurately estimate the noise $d(n)$ in either the time domain or frequency domain.

2.2. Noise Estimation

The noise estimate for performing speech enhancement can be obtained in either the time domain or frequency domain from the noisy observation model

$$y(n) = x(n) + d(n) \tag{2.5}$$

with uncorrelated additive noise $d(n)$. By dividing $y(n)$ into overlapping frames and applying a window function (e.g., Hanning or Hamming), (2.5) can be analyzed in the frequency domain as

$$Y(k,l) = \sum_{n=0}^{N-1} y(n+lM)h(n)e^{-j\left(\frac{2\pi}{N}\right)nk}, \tag{2.6}$$

where k is the frequency bin index, l is the time frame index, h is the analysis window of size N, and M is the frame update step in time. The presence or absence of speech in the l^{th} frame of the k^{th} frequency bin are described by the hypotheses

$$\begin{aligned} H_0(k,l):\ Y(k,l) &= D(k,l) \\ H_1(k,l):\ Y(k,l) &= X(k,l) + D(k,l) \end{aligned}, \tag{2.7}$$

where $X(k,l)$ and $D(k,l)$ represent the short-time Fourier transform (STFT) of the clean signal and noise signal. The spectral noise variance, which is denoted as $\lambda_d(k,l) = E\left[\left|D(k,l)\right|^2\right]$ in the k^{th} frequency bin, is commonly tracked and estimated by applying a temporal recursive smoothing to the noisy observation $Y(k,l)$ during periods of speech absence. In particular, the update equations for speech absence and speech presence are

$$\begin{aligned} H_0'(k,l):\ \hat{\lambda}_d(k,l+1) &= \alpha_d \hat{\lambda}_d(k,l) + (1-\alpha_d)\left|Y(k,l)\right|^2 \\ H_1'(k,l):\ \hat{\lambda}_d(k,l+1) &= \hat{\lambda}_d(k,l) \end{aligned}, \tag{2.8}$$

where α_d is a smoothing parameter defined as $0 < \alpha_d < 1$. Overall, there are several techniques for estimating the spectral noise variance $\lambda_d(k,l)$ that include the simple

silence detection as well as more advanced methods such as Minimum Statistics (MS) [17] and Recursive Averaging [18, 19].

2.2.1. Silence Detection

Silence detection is perhaps the simplest form of estimating the noise statistics for a given noisy speech utterance. Voice activity detectors (VADs) [20] are used to determine periods of speech absence and speech presence based on comparing an extract feature (e.g., short-time energy, zero-crossings) against a particular threshold, which is usually determined during speech absence periods. Based on the labeled speech absence and speech presence regions, the noise spectrum is estimated as

$$\left|\hat{N}\left(k,l\right)\right| = \frac{1}{C\left(I\left(l\right)=1\right)}\sum_{l=0}^{L-1} I\left(l\right)\left|Y\left(k,l\right)\right|, \tag{2.9}$$

where the indicator function $I\left(l\right)$ is defined by

$$I\left(l\right) = \begin{cases} 1, & \text{speech absence} \\ 0, & \text{speech presence} \end{cases} \tag{2.10}$$

with $C\left(I\left(l\right)=1\right)$ representing the count of frames in which $I\left(l\right)=1$. As an even simpler approach, the first L frames of the noisy signal can be assumed to contain silence (i.e., strictly noise) and the noise spectrum is estimated by averaging those initial L frames before the occurrence of speech in the $L+1$ frame.

The reliability of the noise estimate from silence detection methods severely deteriorates for weak speech components, low input SNR, and insufficient non-speech sections. Although the approaches work satisfactorily for stationary noise conditions

(e.g., white noise), silence detectors often fail in more realistic non-stationary noise conditions (e.g., restaurant noise) with constantly changing spectral noise characteristics. Therefore, the better scheme for estimating the noise spectrum requires tracking the noise spectrum continuously over time during speech absent and speech present frames through methods such as Minimum Statistics [17] and Recursive Averaging [18, 19].

2.2.2. Minimum Statistics

Minimum Statistics (MS) [17] is a method for estimating the power spectral density of non-stationary noise that does not use voice activity detection (VAD). Fundamentally, the approach tracks spectral minima in each frequency band without any distinction between speech activity and speech pause. The optimal smoothing parameter for recursive smoothing of the power spectral density of the noisy speech signal is based on minimizing a conditional mean-square error (MSE) estimation criterion in each time step. From the optimally smoothed power spectral density estimate and analysis of the statistics of the spectral minima, an unbiased noise estimator is developed that is well-suited for real time implementation. The smoothing parameter α is

$$\alpha_{opt}(k,l) = \frac{\alpha_{\max}\alpha_c(k)}{1+\left(\frac{P(k,l-1)}{\hat{\lambda}_d(k,l-1)-1}\right)^2} \tag{2.11}$$

with

$$\alpha_c(l) = \frac{1}{1 + \left(\frac{\sum_{k=0}^{M-1} P(k,l-1)}{\sum_{k=0}^{M-1} |Y(k,l)|^2} - 1 \right)^2}, \tag{2.12}$$

and the smoothing power spectrum $P(k,l)$ is

$$P(k,l) = \alpha(k,l) P(k,l) + (1 - \alpha(k,l)) |Y(k,l)|^2, \tag{2.13}$$

where the short-term periodogram $|Y(k,l)|^2$ is calculated for each frame l. From the normalized variance

$$\frac{1}{Q_{eq}(k,l)} \approx \frac{\mathrm{var}\left(\hat{P}(k,l)\right)}{2\hat{\sigma}_d^4(k,l-1)}, \tag{2.14}$$

the bias factor $B_{\min}$ is calculated by

$$B_{\min}(k,l) \approx 1 + (D-1)\frac{2}{\hat{Q}_{eq}(k,l)}, \tag{2.15}$$

where D is the window length in number of frames l to search for the minimum. To determine the minimum, $P_{\min}$ is computed and updated as

$$P_{\min}(k,l) = \min\left\{P_{tmp}(k,l-1), P(k,l)\right\} \tag{2.16}$$

with temporary variable P_{tmp} given as

$$P_{tmp}(k,l) = P(k,l) \tag{2.17}$$

for $\mathrm{mod}\left(\frac{l}{D}\right)$ and otherwise as

$$P_{\min}(k,l) = \min\left\{P_{\min}(k,l-1), P(k,l)\right\} \tag{2.18}$$

with

$$P_{tmp}(k,l) = \min\left\{P_{tmp}(k,l-1), P(k,l)\right\}. \tag{2.19}$$

From (2.15) and $P_{\min}$, the noise power spectrum density is computed and updated as

$$\hat{\sigma}_d^2(k,l) = B_{\min}(k,l) P_{\min}(k,l). \tag{2.20}$$

This estimator has superior ability to preserve weak speech sounds and deliver excellent intelligibility as compared to more traditional noise estimation approaches [17].

2.2.3. Recursive Averaging

Minima Controlled Recursive Averaging (MCRA) [18] is an approach for noise estimation that averages past spectral power values using a smoothing parameter adjusted by signal presence probability in each of the sub-bands. Presence of speech in the sub-bands is determined by the ratio between the local energy of the noisy speech and its minimum within a specified time window. The local energy of the noisy speech is obtained by smoothing the magnitude squared short-time Fourier transform (STFT) in both the time and frequency domains. In the frequency domain, the smoothed magnitude squared STFT is

$$S_f(k,l) = \sum_{i=-w}^{w} b(i)\left|Y(k-i,l)\right|^2, \tag{2.21}$$

where b represents a window function such as Hamming or Hanning with length $2w+1$ and k and l are the frequency bin index and frame. In the time domain, the smoothing is performed by a first-order recursive averaging

$$S(k,l) = \alpha_s S(k,l-1) + (1-\alpha_s) S_f(k,l) \tag{2.22}$$

with parameter $\alpha_s \in (0,1)$. The minimum of the local energy $S_{\min}(k,l)$ is searched by first initializing the minimum $S_{\min}(k,0) = S(k,0)$ and temporary $S_{tmp}(k,0) = S(k,0)$ variables. Then, there is a sample-wise comparison of the local energy and minimum value of the previous frame to produce the minimum value for the current frame as

$$S_{\min}(k,l) = \min\{S_{\min}(k,l-1), S(k,l)\} \tag{2.23}$$

with

$$S_{tmp}(k,l) = \min\{S_{tmp}(k,l-1), S(k,l)\}. \tag{2.24}$$

If l is divisible by the number of read frames L, then (2.23) and (2.24) are rewritten as

$$S_{\min}(k,l) = \min\{S_{tmp}(k,l-1), S(k,l)\} \tag{2.25}$$

with

$$S_{tmp}(k,l) = S(k,l) \tag{2.26}$$

with the search continuing again with (2.23) and (2.24). With the threshold δ, the ratio

$$S_r(k,l) = \frac{S(k,l)}{S_{\min}(k,l)} \tag{2.27}$$

is computed to decide the speech presence regions in the indicator function $I(k,l)$ as

$$I(k,l) = \begin{cases} I(k,l) = 1, \text{ if } S_r(k,l) > \delta \text{ (speech present)} \\ I(k,l) = 0, \text{ if } S_r(k,l) < \delta \text{ (speech absent)} \end{cases}. \tag{2.28}$$

Consequently, the speech presence probability $p(k,l)$ is smoothed over time using a first-order recursive averaging

$$\hat{p}'(k,l) = \alpha_p \hat{p}'(k,l-1) + (1-\alpha_p) I(k,l) \tag{2.29}$$

with parameter $\alpha_p \in (0,1)$. With (2.29), the time-varying smoothing parameter $\tilde{\alpha}_d$ is computed as

$$\tilde{\alpha}_d(k,l) = \alpha_d + (1-\alpha_d) p'(k,l), \tag{2.30}$$

which is then substituted into the noise estimator $\hat{\lambda}_d$

$$\hat{\lambda}_d(k,l+1) = \tilde{\alpha}_d(k,l)\hat{\lambda}_d(k,l) + \left[1-\tilde{\alpha}_d(k,l)\right]\left|Y(k,l)\right|^2. \tag{2.31}$$

The MCRA noise estimation algorithm is a computationally efficient method, robust with respect to the input SNR and SSNR and type of underlying uncorrelated additive noise, and characterized by the ability to quickly track abrupt changes in the noise spectrum for non-stationary noises [18].

Improved Minima Controlled Recursive Averaging (IMCRA) [19] is an extension of the MCRA noise estimation algorithm that involves non-stationary noise, weak speech components, and low input SNR. The noise estimate is obtained by averaging past spectral power values using a time-varying, frequency-dependent smoothing parameter adjusted by the signal presence probability, which is controlled by minima values of a smoothing periodogram. Fundamentally, the IMCRA algorithm consists of two iterations: 1) rough VAD in each frequency band and 2) smoothing that excludes relatively strong

speech components. By using (2.22), the smooth power spectrum $S(k,l)$ is updated in the first iteration as

$$S_{\min}(k,l) = \min\{S_{\min}(k,l-1), S(k,l)\} \tag{2.32}$$

with

$$S_{\min\_sw}(k) = \min\{S_{\min\_sw}(k), S(k,l)\}. \tag{2.33}$$

Based on the threshold parameters

$$\gamma_{\min}(k,l) = \frac{|Y(k,l)|^2}{B_{\min} S_{\min}(k,l)} \tag{2.34}$$

and

$$\varsigma(k,l) = \frac{S(k,l)}{B_{\min} S_{\min}(k,l)}, \tag{2.35}$$

the indicator function $I(k,l)$ is computed as

$$I(k,l) = \begin{cases} 1, \ \gamma_{\min}(k,l) < \gamma_0 \text{ and } \varsigma(k,l) < \varsigma_0 \text{ (speech absent)} \\ 0, \ \text{otherwise (speech present)} \end{cases}. \tag{2.36}$$

For the second iteration, the smoothing power spectrum $\tilde{S}(k,l)$ is computed as in (2.22) with

$$\tilde{S}_f(k,l) = \begin{cases} \dfrac{\sum_{i=-w}^{w} b(i) I(k-i,l) |Y(k-i,l)|^2}{\sum_{i=-w}^{w} b(i) I(k-i,l)}, \ \text{if } \sum_{i=-w}^{w} I(k-i,l) \neq 0 \\ \tilde{S}(k,l-1), \ \text{otherwise} \end{cases} \tag{2.37}$$

and updated as

$$\tilde{S}_{\min}(k,l) = \min\left\{\tilde{S}_{\min}(k,l-1), \tilde{S}(k,l)\right\} \tag{2.38}$$

with

$$\tilde{S}_{\min\_sw}(k) = \min\left\{\tilde{S}_{\min\_sw}(k), \tilde{S}(k,l)\right\}. \tag{2.39}$$

From the threshold parameters

$$\tilde{\gamma}_{\min}(k,l) = \frac{\left|Y(k,l)\right|^2}{B_{\min}\tilde{S}(k,l)} \tag{2.40}$$

and

$$\tilde{\varsigma}(k,l) = \frac{S(k,l)}{B_{\min}\tilde{S}_{\min}(k,l)}, \tag{2.41}$$

the *a priori* speech absence probability estimator is computed by

$$\hat{q}(k,l) = \begin{cases} 1, \text{ if } \tilde{\gamma}_{\min}(k,l) \le 1 \text{ and } \varsigma(k,l) < \varsigma_0 \\ \left(\dfrac{\gamma_1 - \tilde{\gamma}_{\min}}{\gamma_1 - 1}\right), \text{ if } 1 < \tilde{\gamma}_{\min}(k,l) < \gamma_1 \text{ and } \tilde{\varsigma}(k,l) < \varsigma_0 \\ 0, \text{ otherwise} \end{cases}. \tag{2.42}$$

The speech presence probability $p(k,l)$ is calculated as

$$p(k,l) = \left\{1 + \frac{q(k,l)}{1-q(k,l)}\left(1+\xi(k,l)\right)\exp\left(-v(k,l)\right)\right\}^{-1}, \tag{2.43}$$

where $q(k,l) = P(H_0(k,l))$ is the *a priori* probability for speech absence with $\xi = \frac{\sigma_X^2}{\sigma_N^2}$ and $v = \frac{\xi}{1+\xi}\gamma$, where $\gamma = \frac{|Y|^2}{\sigma_N^2}$. In a similar fashion to the MCRA noise estimation algorithm, the time-varying, frequency-dependent smoothing parameter is updated as in (2.30) but with the noise estimator written as

$$\hat{\lambda}_d(k,l+1) = \beta\bar{\lambda}_d(k,l+1), \tag{2.44}$$

where $\bar{\lambda}_d$ is estimated as in (2.31) with bias compensation factor β. In comparison to the MCRA [18] noise estimation technique, the IMCRA algorithm obtains a lower estimation error and improves speech quality and lowers residual noises for speech enhancement [19].

2.3. Single Channel Enhancement

With the estimate of the noise statistics, the estimate of the clean speech signal $\hat{s}(t)$ can be obtained through several time domain or frequency domain methods. In the next section, traditional single channel speech enhancement methods are explained in detail.

2.3.1. Spectral Subtraction

Spectral subtraction [21] is a noise suppression algorithm that reduces the spectral effects of acoustically-added noise in speech. The basic assumption is that the desired clean signal $s(t)$ has been corrupted by uncorrelated additive noise $n(t)$ to produce a noisy signal as in (2.2). In the frequency domain, (2.2) is written as

$$Y(k,l) = X(k,l) + N(k,l). \tag{2.45}$$

From (2.45), the power spectrum is computed as

$$|Y(k,l)|^2 = |X(k,l)|^2 + |N(k,l)|^2 + 2|X(k,l)||N(k,l)|\cos(\theta(k,l)), \tag{2.46}$$

where $\cos(\theta(k,l))$ is the random angle between the two complex variables for the speech $X(k,l)$ and noise $N(k,l)$. By assuming that $X(k,l)$ and $N(k,l)$ are orthogonal to each other, $\cos(\theta(k,l)) = 0$ and (2.46) is rewritten as

$$|Y(k,l)|^2 = |X(k,l)|^2 + |N(k,l)|^2 \tag{2.47}$$

or

$$|X(k,l)|^2 = |Y(k,l)|^2 - |N(k,l)|^2. \tag{2.48}$$

While $|Y(k,l)|^2$ can be directly computed from the given noisy observation $Y(k,l)$, $N(k,l)$ must be determined by means of a noise estimation technique. The clean speech signal is defined in the frequency domain as

$$\begin{aligned} X(k,l) &= |X(k,l)|e^{j\angle X(k,l)} \\ &= |X(k,l)|e^{j\angle \alpha(k,l)}, \end{aligned} \tag{2.49}$$

which needs an estimate of the spectral amplitude $|X(k,l)|$ and spectral phase $\alpha(k,l)$. Similarly, the noisy speech signal is defined in the frequency domain as

$$\begin{aligned} Y(k,l) &= |Y(k,l)|e^{j\angle Y(k,l)} \\ &= |Y(k,l)|e^{j\angle \vartheta(k,l)}. \end{aligned} \tag{2.50}$$

With the noisy signal spectral phase $\vartheta(k,l)$ serving as the estimate of the clean signal spectral phase $\hat{\alpha}(k,l)$ and squared spectral amplitude in (2.48), the clean signal $X(k,l)$ in (2.49) is estimated as

$$\hat{X}(k,l) = \left[|Y(k,l)|^2 - |N(k,l)|^2 \right]^{\frac{1}{2}} e^{j\vartheta(k,l)}. \tag{2.51}$$

From (2.51), the clean signal estimate $x(t)$ is computed by using the inverse short-time Fourier transform (I-STFT). Figure 2-2 illustrates the spectral subtraction technique.

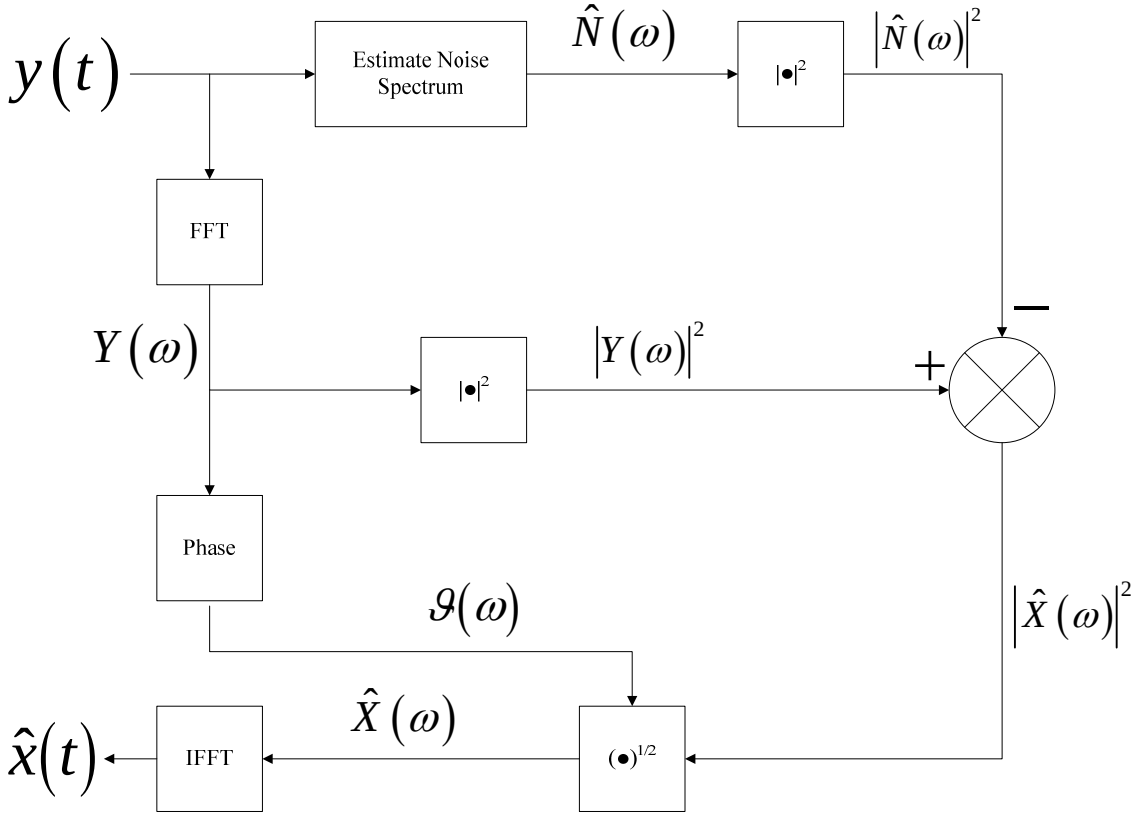

Figure 2-2 Spectral Subtraction

Spectral subtraction is a computationally simple signal enhancement algorithm that exhibits significant noise reduction but experiences warbling noise with tonal quality

referred to as musical noise [22], which results from rapid changes with frequency [21] and misestimation of the noise spectrum [20] primarily in unvoiced segments of the speech with similar levels of noise power and speech power.

2.3.2. Wiener Filter

Wiener filtering [11] is the optimal minimum mean-square error (MMSE) linear filter for suppressing additive noise in a noisy signal in either the time domain or frequency domain [23]. In the time domain, the estimation error $e(n)$ is computed as the difference between the desired signal $d(n) = x(n)$ and estimated desired signal $\hat{d}(n)$ as

$$\begin{aligned} e(n) &= d(n) - \hat{d}(n) \\ &= d(n) - \sum_{k=0}^{M-1} h_k y(n-k), \\ &= d(n) - h^T y \end{aligned} \tag{2.52}$$

where $h^T = [h_0, h_1, ..., h_{M-1}]$ is the finite impulse response (FIR) filter coefficient vector and $y^T = [y(n), y(n-1), ..., y(n-M+1)]$ is the input vector containing the past M samples of the input for $n = 0, 1, 2, \ldots$. To determine the optimal filter coefficients, the MSE of (2.52) is written as

$$J = E[e^2(n)] = E[d^2(n)] - 2h^T r_{yd}^{-} + h^T R_{yy} h \tag{2.53}$$

and minimized as

$$\frac{\partial J}{\partial h_k} = 0 = 2E\left[e(n) \frac{\partial e(n)}{\partial h_k} \right] \tag{2.54}$$

for $k = 0, 1, ..., M-1$. From (2.52), the partial derivative for the error $e(n)$ in (2.54) is

$$\frac{\partial e(n)}{\partial h_k} = -y(n-k). \tag{2.55}$$

By substitution of (2.55) into (2.54),

$$\frac{\partial J}{\partial h_k} = -2E\left[e(n)\,y(n-k)\right] = 0, \tag{2.56}$$

which is the orthogonality principle of optimum linear filtering. In vector and matrix notation, (2.56) is written as

$$\frac{\partial J}{\partial h} = -2r_{yd}^{-} + 2h^T R_{yy} = 0 \tag{2.57}$$

with

$$R_{yy}h^{*} = r_{yd}^{-} \tag{2.58}$$

or

$$h^{*} = R_{yy}^{-1}r_{yd}^{-}, \tag{2.59}$$

which are the Wiener-Hopf solutions [24]. In (2.59), $r_{yd}^{-} = E\left[yd(n)\right] = E\left[\left(y(n)\,y(n-1)\cdots y(n-M+1)\right)d(n)\right]$ is defined as the cross-correlation vector $(M\times 1)$ between the input and desired signals and $R_{yy} = E\left[yy^T\right]$ is the autocorrelation matrix $(M\times M)$ of the input signal. To evaluate the time domain Wiener filter, R_{yy} and r_{yd}^{-} must be computed in (2.59). By definition,

$$R_{yy} = E\left[yy^T\right] = R_{xx} + R_{nn}, \tag{2.60}$$

where the last expectation two terms in (2.60) are zero since the speech and noise are assumed uncorrelated and zero-mean. By using the assumption of uncorrelated speech and noise, the cross-correlation vector r_{yd}^{-} is

$$r_{yd}^{-} = E\left[yd\left(n\right)\right] = r_{xx} \tag{2.61}$$

Through (2.60) and (2.61), the Wiener filter in the time domain for (2.59) is rewritten as

$$\begin{aligned} h^{*} &= R_{yy}^{-1} r_{yd}^{-} \\ &= \left(R_{xx} + R_{nn}\right)^{-1} r_{xx} \end{aligned}. \tag{2.62}$$

As an alternative derivation to the time domain FIR Wiener filter in (2.62), the Wiener filter can be derived in the frequency domain as a two-sided, infinite impulse response (IIR) filter. In the frequency domain, the estimate of the desired response $D\left(\omega_k\right) = X\left(\omega_k\right)$ is

$$\hat{D}\left(\omega_k\right) = H\left(\omega_k\right) Y\left(\omega_k\right) \tag{2.63}$$

with estimation error

$$\begin{aligned} E\left(\omega_k\right) &= D\left(\omega_k\right) - \hat{D}\left(\omega_k\right) \\ &= D\left(\omega_k\right) - H\left(\omega_k\right) Y\left(\omega_k\right) \end{aligned}, \tag{2.64}$$

where $H\left(\omega_k\right)$ is the gain function that is applied to the noisy observation $Y\left(\omega_k\right)$. To compute $H\left(\omega_k\right)$, the MSE of (2.64) is defined as

$$\begin{aligned} J &= E\left[\left|E\left(\omega_k\right)\right|^2\right] \\ &= E\left[\left|D\left(\omega_k\right)\right|^2\right] - H^{*}\left(\omega_k\right) P_{dy}\left(\omega_k\right) - H\left(\omega_k\right) P_{yd}\left(\omega_k\right) + \left|H\left(\omega_k\right)\right|^2 P_{yy}\left(\omega_k\right) \end{aligned} \tag{2.65}$$

and minimized with respect to $H(\omega_k)$ as

$$\frac{\partial J}{\partial H(\omega_k)} = 0 = -P_{yd}(\omega_k) + H^*(\omega_k) P_{yy}(\omega_k), \tag{2.66}$$

where

$$P_{yd}(\omega_k) = E\left[Y(\omega_k) D^*(\omega_k)\right] = P_{xx}(\omega_k) \tag{2.67}$$

and

$$P_{dy}(\omega_k) = E\left[D(\omega_k) Y^*(\omega_k)\right] = P_{yd}(\omega_k) \tag{2.68}$$

and

$$P_{yy}(\omega_k) = E\left[Y(\omega_k) Y^*(\omega_k)\right] = P_{xx}(\omega_k) + P_{nn}(\omega_k) \tag{2.69}$$

with $P_{\bullet\bullet}(\omega_k)$ serving as the power spectrums. By solving (2.66) for the gain function $H^*(\omega_k)$ and substituting (2.68) and (2.69), the Wiener filter in the frequency domain is written as

$$H^*(\omega_k) = \frac{P_{yd}(\omega_k)}{P_{yy}(\omega_k)} = \frac{P_{xx}(\omega_k)}{P_{xx}(\omega_k) + P_{nn}(\omega_k)} \tag{2.70}$$

or

$$H^*(\omega_k) = \frac{\xi(\omega_k)}{1 + \xi(\omega_k)}, \tag{2.71}$$

where $\xi(\omega_k)$ is the *a priori* SNR defined as

$$\xi(\omega_k) = \frac{P_{xx}(\omega_k)}{P_{nn}(\omega_k)}. \tag{2.72}$$

As a note, the Wiener filter in (2.70) is IIR and non-causal, which means that it is not realizable in its current form. In comparison to spectral subtraction method, which requires only an estimate of the noise power spectrum $P_{xx}(\omega_k)$, Wiener filtering requires both an estimate of the power spectrum of both the clean speech power spectrum $P_{xx}(\omega_k)$ and noise power spectrum $P_{nn}(\omega_k)$. Since the Wiener filter in (2.70) requires an estimate of the unknown $P_{xx}(\omega_k)$, (2.70) can be reformulated in an iterative form. Figure 2-3 illustrates the iterative Wiener filtering process.

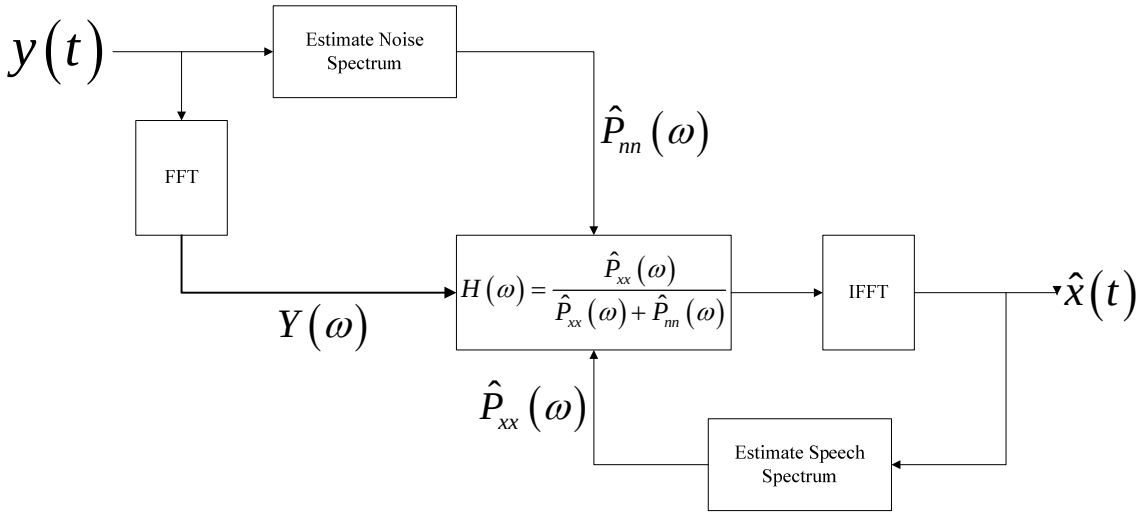

Figure 2-3 Frequency Domain Iterative Wiener Filter

As with spectral subtraction, Wiener filters in both the time domain (2.62) and frequency domain (2.71) still suffer from musical tones in the enhanced speech signal but show improvements in SNR and SSNR performance.

2.3.3. Short-Time Spectral Amplitude Estimation

As opposed to the optimal linear MMSE estimator, the non-linear MMSE short-time spectral amplitude (STSA) estimator [5] can be derived for the clean speech spectrum. From (2.45), the noisy observation $Y(k,l)$ can be expressed in terms of its spectral amplitude and spectral phase as

$$\begin{aligned} Y(k,l) &= X(k,l) + N(k,l) \\ |Y(k,l)|e^{j\angle Y(k,l)} &= |X(k,l)|e^{j\angle X(k,l)} + N(k,l) \\ R(k,l)e^{j\angle \vartheta(k,l)} &= A(k,l)e^{j\angle \alpha(k,l)} + N(k,l) \end{aligned} \tag{2.73}$$

or

$$\mathrm{Re}^{j\angle \vartheta} = Ae^{j\angle \alpha} + N \tag{2.74}$$

without explicit dependencies k and l. By minimizing the MSE between the true spectral amplitude A and estimated true spectral amplitude $\hat{A}$ and using Bayes rule, the MMSE STSA estimator is written as

$$\hat{A}_{STSA} = E\left[A|Y\right] = \frac{\int_0^{\infty}\int_0^{2\pi} Ap(Y|A,\alpha)\,p(A,\alpha)\,d\alpha dA}{\int_0^{\infty}\int_0^{2\pi} p(Y|A,\alpha)\,p(A,\alpha)\,d\alpha dA} \tag{2.75}$$

with speech prior

$$p(A,\alpha) = \frac{A}{\pi\sigma_x^2}\exp\left(-\frac{A^2}{\sigma_x^2}\right) \tag{2.76}$$

and noise likelihood

$$p\left(Y\middle|A,\alpha\right)=\frac{1}{\pi\sigma_n^2}\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right) \tag{2.77}$$

as the assumed statistical models with statistical independence between the spectral components $Y\left(k,l\right)$. After substitution of the statistical models in (2.76) and (2.77), the result from (2.75) is

$$\hat{A}_{STSA}=\frac{\int_0^\infty A^2\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha dA}{\int_0^\infty A\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha dA}. \tag{2.78}$$

The integration over the spectral phase α is performed by expansion of the term $\left|Y-Ae^{j\alpha}\right|^2=\left(Y-Ae^{j\alpha}\right)_R^2+\left(Y-Ae^{j\alpha}\right)_I^2$ and extracting the constants from the integral as

$$\begin{aligned}&\int_0^{2\pi}\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha\\&=\exp\left(-\frac{\left|Y\right|^2+A^2}{\sigma_n^2}\right)\int_0^{2\pi}\exp\left(a\cos\alpha+b\sin\alpha\right)d\alpha\end{aligned}, \tag{2.79}$$

where

$$a=\frac{2A}{\sigma_n^2}\mathrm{Re}\left(Y\right) \tag{2.80}$$

and

$$b=\frac{2A}{\sigma_n^2}\mathrm{Im}\left(Y\right). \tag{2.81}$$

From trigonometric identities, the sum of cosine and sine terms with different amplitudes and the same phase is written as

$$a\cos\alpha + b\sin\alpha = \sqrt{a^2 + b^2}\cos\left(\alpha - \arctan\left(\frac{b}{a}\right)\right), \tag{2.82}$$

where

$$\sqrt{a^2 + b^2} = 2A\left|\frac{Y}{\sigma_n^2}\right|. \tag{2.83}$$

Since the integral in (2.79) for the spectral phase α is over one full period, the spectral phase shift of $\arctan\left(\frac{b}{a}\right)$ is removed from (2.82). By means of equation 8.431.1 in [25], the integral in (2.79) is rewritten as

$$\int_0^{2\pi} \exp\left(a\cos\alpha + b\sin\alpha\right) d\alpha = 2\pi I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right), \tag{2.84}$$

where $I_0(x)$ is the modified Bessel function of the first kind of order 0, which reduces (2.78) to the form

$$\hat{A}_{STSA} = \frac{\int_0^{\infty} A^2 \exp\left(-A^2\frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA}{\int_0^{\infty} A\exp\left(-A^2\frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA}. \tag{2.85}$$

Through substitution of equations 8.406.3 and 6.631.1 in [26] and [25], the estimator is expressed as

$$\begin{aligned}\hat{A}_{STSA} &= \Gamma(1.5)\frac{\sqrt{v}}{\gamma}\,{}_1F_1(-0.5;1;-v)R \\ &= \Gamma(1.5)\frac{\sqrt{v}}{\gamma}\exp\left(-\frac{v}{2}\right)\left[(1+v)I_0\left(\frac{v}{2}\right)+vI_1\left(\frac{v}{2}\right)\right]R\end{aligned} \tag{2.86}$$

with

$$v = \frac{\xi}{1+\xi}\gamma \tag{2.87}$$

and

$$\xi = \frac{\sigma_x^2}{\sigma_n^2} \tag{2.88}$$

and

$$\gamma = \frac{R^2}{\sigma_n^2}, \tag{2.89}$$

where ξ and γ are the *a priori* and *a posteriori* SNR and $\Gamma(\bullet)$ and ${}_1F_1(\bullet;\bullet;\bullet)$ denote the gamma function and confluent hypergeometric function as described by equation 9.210 in [25]. The gain function G_{STSA} is defined as

$$G_{STSA} = \frac{\hat{A}_{STSA}}{R} = \Gamma(1.5)\frac{\sqrt{v}}{\gamma}\exp\left(-\frac{v}{2}\right)\left[(1+v)I_0\left(\frac{v}{2}\right)+vI_1\left(\frac{v}{2}\right)\right] \tag{2.90}$$

or

$$\hat{A}_{STSA} = G_{STSA}R, \tag{2.91}$$

which reformulates the optimal MMSE STSA estimator as a filter similar to the Wiener filter. Based on the spectral amplitude estimator $\hat{A}$, there is a significant reduction of

noise and enhanced speech with colorless noise compared to the spectral subtraction and Wiener filter methods [5].

In the STSA estimator in (2.86), $\hat{A}_{STSA}$ requires estimation of the spectral noise variance σ_n^2 and spectral speech variance σ_x^2. Whereas the spectral noise variance σ_n^2 can be estimated using various noise estimation techniques, the spectral speech variance σ_x^2 can be estimated using a maximum likelihood (ML) estimate or decision-directed approach (DD) [5]. By maximizing the joint conditional

$$\begin{aligned} & p\left(Y\left(k,n\right)\middle|\sigma_x^2\left(k\right),\sigma_n^2\left(k\right)\right) \\ & =\prod_{l=0}^{L-1}\frac{1}{\pi\left(\sigma_x^2\left(k\right)+\sigma_n^2\left(k\right)\right)}\exp\left(-\frac{R^2\left(k,n-l\right)}{\sigma_x^2\left(k\right)+\sigma_n^2\left(k\right)}\right), \end{aligned} \tag{2.92}$$

the ML spectral speech variance estimator in the l^{th} analysis frame is

$$\hat{\sigma}_x^2\left(k\right)=\begin{cases}\dfrac{1}{L}\displaystyle\sum_{l=0}^{L-1}R^2\left(k,n-l\right)-\sigma_n^2\left(k\right), & \text{if non-negative}\\ 0, & \text{otherwise}\end{cases}. \tag{2.93}$$

From (2.88) and (2.93), the *a priori* SNR ξ is written as

$$\hat{\xi}\left(k\right)=\begin{cases}\dfrac{1}{L}\displaystyle\sum_{l=0}^{L-1}\gamma\left(k,n-l\right)-1, & \text{if non-negative}\\ 0, & \text{otherwise}\end{cases}. \tag{2.94}$$

In contrast, the derivation of the DD approach is based on the definition of the *a priori* SNR ξ

$$\xi\left(k,n\right)=\frac{E\left[A^2\left(k,n\right)\right]}{\sigma_n^2\left(k,n\right)} \tag{2.95}$$

and its relation to the *a posteriori* SNR γ

$$\xi(k,n) = E\left[\gamma(k,n) - 1\right]. \tag{2.96}$$

From (2.95) and (2.96), ξ is written as

$$\xi(k,n) = E\left[\frac{1}{2}\frac{A^2(k,n-1)}{\sigma_n^2(k,n-1)} + \frac{1}{2}\left(\gamma(k,n) - 1\right)\right]. \tag{2.97}$$

The DD estimator $\hat{\xi}$ is deduced from (2.97) as

$$\hat{\xi}(k,n) = \alpha\frac{\hat{A}^2(k,n-1)}{\sigma_n^2(k,n-1)} + (1-\alpha)P\left[\gamma(k,n) - 1\right], \tag{2.98}$$

where α is a smoothing parameter defined as $0 \le \alpha \le 1$ and $P[\bullet]$ is the operator

$$P[x] = \begin{cases} x, & \text{if } x \ge 0 \\ 0, & \text{otherwise} \end{cases}. \tag{2.99}$$

By combining the DD estimator in (2.98) with the MMSE STSA estimator in (2.86), the best speech enhancement results are obtained from the noisy observations.

2.3.4. Log-Spectral Amplitude Estimation

Similarly to the MMSE STSA, the non-linear log-spectral amplitude estimator (LSA) is also an MMSE spectral amplitude estimator for speech enhancement [5, 27]. While the MMSE STSA estimator minimizes the squared error between the spectral amplitude and estimated spectral amplitude, the MMSE LSA minimizes the squared error between the log-spectral amplitude and estimated log-spectral amplitude, which is a more

subjectively meaningful distortion measure that correlates well with human perception [6]. The MMSE LSA estimator is written as

$$\begin{aligned}\hat{A}_{LSA} &= \exp\left(E\left[\ln(A)|Y\right]\right) \\ &= \exp\left(E\left[Z|Y\right]\right)\end{aligned}. \tag{2.100}$$

In the log-spectral amplitude estimator in (2.100), the expectation $E\left[Z|Y\right]$ is evaluated by

$$E\left[Z|Y\right] = \frac{d}{d\mu}\left[\Phi_{Z|Y}(\mu)\right]\Big|_{\mu=0}, \tag{2.101}$$

where $\Phi_{Z|Y}(\mu) = E\left[A^{\mu}|Y\right]$ is defined as the moment-generating function

$$\Phi_{Z|Y}(\mu) = \frac{\int_0^{\infty}\int_0^{2\pi} A^{\mu}\, p\left(Y|A,\alpha\right) p\left(A,\alpha\right) d\alpha dA}{\int_0^{\infty}\int_0^{2\pi} p\left(Y|A,\alpha\right) p\left(A,\alpha\right) d\alpha dA}. \tag{2.102}$$

After substitution of (2.76) and (2.77), (2.102) is expressed as

$$\Phi_{Z|Y}(\mu) = \frac{\int_0^{\infty} A^{\mu+1}\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|^2}{\sigma_n^2}\right) d\alpha dA}{\int_0^{\infty} A\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|^2}{\sigma_n^2}\right) d\alpha dA}. \tag{2.103}$$

The integration over the spectral phase α is performed exactly as for the spectral amplitude estimator. By employing (2.79)-(2.84), (2.103) is written as

$$\Phi_{Z|Y}(\mu) = \frac{\int_0^\infty A^{\mu+1} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA}{\int_0^\infty A \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA}, \tag{2.104}$$

where

$$\frac{1}{\lambda} = \frac{1}{\sigma_x^2} + \frac{1}{\sigma_n^2}. \tag{2.105}$$

Through application of equations 8.406.3 and 6.631.1 in [26] and [25], $\Phi_{Z|Y}(\mu)$ is expressed as

$$\Phi_{Z|Y}(\mu) = \frac{\Gamma\left(\frac{\mu}{2}+1\right)}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}} {}_1F_1\left(-\frac{\mu}{2};1;-v\right), \tag{2.106}$$

where

$$v = \frac{\left|\frac{Y}{\sigma_n^2}^2\right|}{\frac{1}{\lambda}} \tag{2.107}$$

and $\Gamma(\bullet)$ and ${}_1F_1(\bullet;\bullet;\bullet)$ denote the gamma function and confluent hypergeometric function as described by equation 9.210 in [25].

The differentiation of (2.106) with respect to μ results in three derivative terms that are written as

$$
\begin{aligned}
E\left[Z|Y\right] &= \frac{d}{d\mu}\left[\Phi_{Z|Y}\left(\mu\right)\right]\Big|_{\mu=0} \\
&= \left[\frac{d}{d\mu}\left(\Gamma\left(\frac{\mu}{2}+1\right)\right)\frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}}{}_1F_1\left(-\frac{\mu}{2};1;-\nu\right)\right]\Bigg|_{\mu=0} \\
&+ \left[\Gamma\left(\frac{\mu}{2}+1\right)\frac{d}{d\mu}\left(\frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}}\right){}_1F_1\left(-\frac{\mu}{2};1;-\nu\right)\right]\Bigg|_{\mu=0} \\
&+ \left[\Gamma\left(\frac{\mu}{2}+1\right)\frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}}\frac{d}{d\mu}\left({}_1F_1\left(-\frac{\mu}{2};1;-\nu\right)\right)\right]\Bigg|_{\mu=0}
\end{aligned}
\tag{2.108}
$$

and evaluated at $\mu = 0$. The derivative of the first term is evaluated exactly as in [6] using

$$
\frac{d}{d\mu}\left(\Gamma\left(\frac{\mu}{2}+1\right)\right) = \Gamma\left(\frac{\mu}{2}+1\right)\frac{d}{d\mu}\left(\ln\left(\Gamma\left(\frac{\mu}{2}+1\right)\right)\right). \tag{2.109}
$$

Through the series expansion given by equation 8.342.1 in [26], the last term in (2.109) is rewritten as

$$
\ln\left(\Gamma\left(\frac{\mu}{2}+1\right)\right) = -c\frac{\mu}{2} + \sum_{r=2}^{\infty}\frac{\left(-\mu\right)^r}{2^r r}\alpha_r\ , \tag{2.110}
$$

where $\left|\mu\right| < 2$, c is Euler's constant, and

$$
\alpha_r \triangleq \sum_{n=1}^{\infty}\frac{1}{n^r}\,. \tag{2.111}
$$

By differentiating (2.110) term-by-term and evaluating (2.109) at $\mu = 0$, the derivative of the first term in (2.108) is

$$\frac{d}{d\mu}\left[\Gamma\left(\frac{\mu}{2}+1\right)\right]\Bigg|_{\mu=0} = -\frac{c}{2}. \tag{2.112}$$

The derivative of the second term $\left(\frac{1}{\lambda}\right)^{-\frac{\mu}{2}}$ in (2.108) is computed in a straightforward manner by rewriting it in exponential form and evaluating at $\mu = 0$ as

$$\frac{d}{d\mu}\left[\frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}}\right]\Bigg|_{\mu=0} = \frac{d}{d\mu}\left[e^{\frac{1}{2}\mu\ln(\lambda)}\right]\Bigg|_{\mu=0} = \frac{1}{2}\ln(\lambda). \tag{2.113}$$

For the computation of the third term, the confluent hypergeometric function ${}_1F_1\left(-\frac{\mu}{2};1;-\nu\right)$ is differentiated through its series expansion from equation 9.210.1 in [25] as

$$\frac{d}{d\mu}\left[{}_1F_1\left(-\frac{\mu}{2};1;-\nu\right)\right]\Bigg|_{\mu=0} = -\frac{1}{2}\sum_{r=1}^{\infty}\frac{(-\nu)^r}{r!}\frac{1}{r}, \tag{2.114}$$

where $(a)_r = 1\cdot a\cdot(a+1)\cdot\ldots\cdot(a+r-1)$ with $(a)_0 \triangleq 1$. By differentiating (2.114) term-by-term and evaluating at $\mu = 0$, the derivative is

$$\frac{d}{d\mu}\left[{}_1F_1\left(-\frac{\mu}{2};1;-\nu\right)\right]\Bigg|_{\mu=0} = -\frac{1}{2}\sum_{r=1}^{\infty}\frac{(-\nu)^r}{r!}\frac{1}{r}. \tag{2.115}$$

By combining the three derivative results in (2.112), (2.113), and (2.115), (2.108) reduces to

$$\begin{aligned} E\left[Z|Y\right] &= \frac{d}{d\mu}\left[\Phi_{Z|Y}\left(\mu\right)\right]\Big\|_{\mu=0} \\ &= \left(-\frac{c}{2}\right){}_1F_1\left(0;1;-v\right) + \ln\left(\sqrt{\lambda}\right){}_1F_1\left(0;1;-v\right) + \left(-\frac{1}{2}\sum_{r=1}^{\infty}\frac{\left(-v\right)^r}{r!}\frac{1}{r}\right), \\ &= -\frac{1}{2}\left[c + \sum_{r=1}^{\infty}\frac{\left(-v\right)^r}{r!}\frac{1}{r}\right] + \frac{1}{2}\ln\left(\lambda\right) \end{aligned} \tag{2.116}$$

where ${}_1F_1\left(0;1;-v\right) = 1$. From equations 8.211.1 and 8.214.1 in [26], (2.116) is rewritten as

$$\begin{aligned} E\left[Z|Y\right] &= -\frac{1}{2}\left[-\int_v^{\infty}\frac{e^{-t}}{t}dt - \ln\left(v\right)\right] + \frac{1}{2}\ln\left(\lambda\right) \\ &= -\frac{1}{2}\ln\left(\frac{1}{\lambda}\right) + \frac{1}{2}\left[\int_v^{\infty}\frac{e^{-t}}{t}dt + \ln\left(v\right)\right] \end{aligned}. \tag{2.117}$$

From (2.100), the estimator is expressed as

$$\hat{A}_{LSA} = \frac{\xi}{1+\xi}\exp\left(\frac{1}{2}\int_v^{\infty}\frac{e^{-t}}{t}dt\right)R \tag{2.118}$$

or

$$G_{LSA} = \frac{\hat{A}_{LSA}}{R} = \frac{\xi}{1+\xi}\exp\left(\frac{1}{2}\int_v^{\infty}\frac{e^{-t}}{t}dt\right) \tag{2.119}$$

with the *a priori* SNR ξ and v defined in (2.88) and (2.87). Figure 2-4 shows the block diagram of computing the enhanced clean speech estimate $\hat{s}\left(t\right)$.

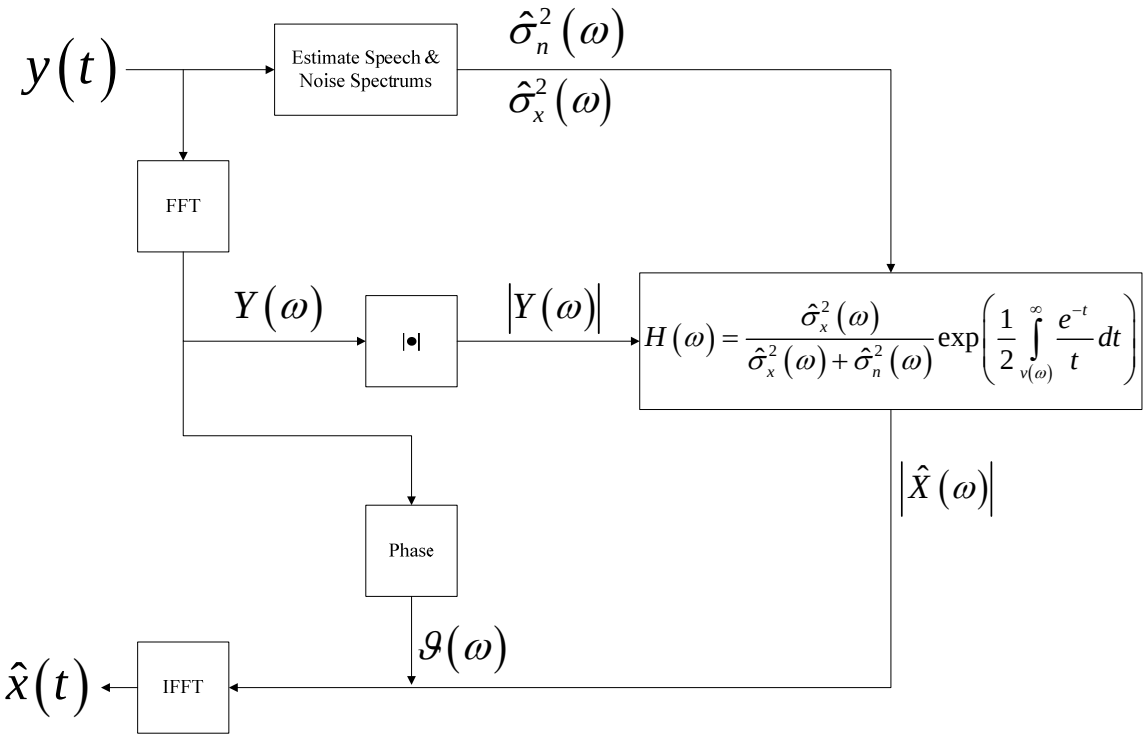

Figure 2-4 Log-Spectral Amplitude (LSA) Estimation

In general, the optimal MMSE LSA estimator $\hat{A}_{LSA}$ is the standard single channel baseline method for comparison against other methods and provides significant reduction of noise, yields enhanced speech with colorless residual noise, and greatly suppresses the musical tone side effects [6].

2.3.5. Perceptually-Motivated Spectral Amplitude Estimation

Bayesian risk functions can be introduced to produce a variety of different spectral amplitude estimators [20]. The Bayes risk is represented as

$$\begin{aligned}\Re_B &= E\left[d\left(A,\hat{A}\right)\right] \\ &= \int\int d\left(A,\hat{A}\right)p\left(A,Y\right)dAdY \\ &= \int p\left(Y\right)\left[\int d\left(A,\hat{A}\right)p\left(A|Y\right)dA\right]dY\end{aligned}, \tag{2.120}$$

where the minimization of the inner integral in (2.120) with respect to the spectral amplitude estimate $\hat{A}$ results in an estimator for each particular cost function. For the STSA cost function [5]

$$d_{STSA}\left(A,\hat{A}\right) = \left(A-\hat{A}\right)^2 \tag{2.121}$$

and LSA cost function [6]

$$d_{LSA}\left(A,\hat{A}\right) = \left(\ln\left(A\right)-\ln\left(\hat{A}\right)\right)^2, \tag{2.122}$$

the resulting estimators are

$$\hat{A}_{STSA} = E\left[A|Y\right] \tag{2.123}$$

and

$$\hat{A}_{LSA} = \exp\left(E\left[\ln\left(A\right)|Y\right]\right). \tag{2.124}$$

In order to incorporate cost functions with perceptual weighting in the estimator, the weighted Euclidean (WE)

$$d_{WE}\left(A,\hat{A}\right) = \left(A-\hat{A}\right)^2 A^p \tag{2.125}$$

and weighted cosh (WCOSH)

$$
\begin{aligned}
d_{WCOSH}\left(A,\hat{A}\right) &= \left[\frac{1}{2}\left(\frac{A}{\hat{A}}+\frac{\hat{A}}{A}\right)-1\right]A^{p} \\
&= \left[\cosh\left(\ln\left(\frac{A}{\hat{A}}\right)\right)-1\right]A^{p} \\
&= \left[\cosh\left(\ln\left(A\right)-\ln\left(\hat{A}\right)\right)-1\right]A^{p}
\end{aligned}
\tag{2.126}
$$

cost functions can be used in (2.120) to construct the optimal MMSE spectral amplitude estimators as

$$
\hat{A}_{WE} = \frac{\int_{0}^{\infty}\int_{0}^{2\pi} A^{p+1}\, p\left(Y\middle|A,\alpha\right) p\left(A,\alpha\right) d\alpha dA}{\int_{0}^{\infty}\int_{0}^{2\pi} A^{p}\, p\left(Y\middle|A,\alpha\right) p\left(A,\alpha\right) d\alpha dA}
\tag{2.127}
$$

and

$$
\hat{A}^{2}_{WCOSH} = \frac{\int_{0}^{\infty}\int_{0}^{2\pi} A^{p+1}\, p\left(Y\middle|A,\alpha\right) p\left(A,\alpha\right) d\alpha dA}{\int_{0}^{\infty}\int_{0}^{2\pi} A^{p-1}\, p\left(Y\middle|A,\alpha\right) p\left(A,\alpha\right) d\alpha dA}
\tag{2.128}
$$

with parameter p. By substitution of the statistical models in (2.76) and (2.77), the results from (2.127) and (2.128) are

$$
\hat{A}_{WE} = \frac{\int_{0}^{\infty} A^{p+2} \exp\left(-\frac{A^{2}}{\sigma_{x}^{2}}\right) \int_{0}^{2\pi} \exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|^{2}}{\sigma_{n}^{2}}\right) d\alpha dA}{\int_{0}^{\infty} A^{p+1} \exp\left(-\frac{A^{2}}{\sigma_{x}^{2}}\right) \int_{0}^{2\pi} \exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|^{2}}{\sigma_{n}^{2}}\right) d\alpha dA}
\tag{2.129}
$$

and

$$\hat{A}^2_{WCOSH} = \frac{\int_0^\infty A^{p+2} \exp\left(-\frac{A^2}{\sigma_x^2}\right) \int_0^{2\pi} \exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|^2}{\sigma_n^2}\right) d\alpha dA}{\int_0^\infty A^{p} \exp\left(-\frac{A^2}{\sigma_x^2}\right) \int_0^{2\pi} \exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|^2}{\sigma_n^2}\right) d\alpha dA}. \tag{2.130}$$

The integration over the spectral phase α is performed exactly as for the STSA and LSA estimators. By employing (2.79)-(2.84), (2.129) and (2.130) are written as

$$\hat{A}_{WE} = \frac{\int_0^\infty A^{p+2} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA}{\int_0^\infty A^{p+1} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA} \tag{2.131}$$

and

$$\hat{A}^2_{WCOSH} = \frac{\int_0^\infty A^{p+2} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA}{\int_0^\infty A^{p} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right) dA}, \tag{2.132}$$

where $\frac{1}{\lambda}$ is defined in (2.105). By utilizing 8.406.3 and 6.631.1 in [26] and [25], (2.131) and (2.132) are written in terms of the gamma function $\Gamma(\bullet)$ and confluent hypergeometric function ${}_1F_1(\bullet;\bullet;\bullet)$ described by 9.210 in [25] as

$$\hat{A}_{WE} = \frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+1\right)} \frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \frac{{}_1F_1\left(\frac{p+3}{2};1;z\right)}{{}_1F_1\left(\frac{p+2}{2};1;z\right)} \tag{2.133}$$

and

$$\hat{A}_{WCOSH}^2 = \frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+\frac{1}{2}\right)}\frac{1}{\frac{1}{\lambda}}\frac{{}_1F_1\left(\frac{p+3}{2};1;z\right)}{{}_1F_1\left(\frac{p+1}{2};1;z\right)}, \tag{2.134}$$

where

$$\frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} = \left(\frac{\sigma_x^2}{1+\xi}\right)^{\frac{1}{2}} \tag{2.135}$$

and

$$\frac{1}{\frac{1}{\lambda}} = \frac{\sigma_x^2}{1+\xi}. \tag{2.136}$$

By using (2.87)-(2.89) in (2.135) and (2.136), the estimators in (2.133) and (2.134) can be expressed as

$$\hat{A}_{WE} = \frac{\sqrt{v}}{\gamma}\frac{\Gamma\left(\frac{p+1}{2}+1\right)}{\Gamma\left(\frac{p}{2}+1\right)}\frac{\Phi\left(-\frac{p+1}{2};1;-v\right)}{\Phi\left(-\frac{p}{2};1;-v\right)}R \tag{2.137}$$

and

$$\hat{A}_{WCOSH} = \frac{1}{\gamma}\sqrt{v\frac{\Gamma\left(\frac{p+3}{2}\right)}{\Gamma\left(\frac{p+1}{2}\right)}\frac{\Phi\left(-\frac{p+1}{2};1;-v\right)}{\Phi\left(-\frac{p-1}{2};1;-v\right)}}R. \tag{2.138}$$

Figure 2-5 shows comparisons of the optimal MMSE perceptually-motivated cost function estimators in (2.137) and (2.138) against the traditional STSA and LSA

estimators in (2.86) and (2.118) based on SSNR improvement in a speech enhancement task.

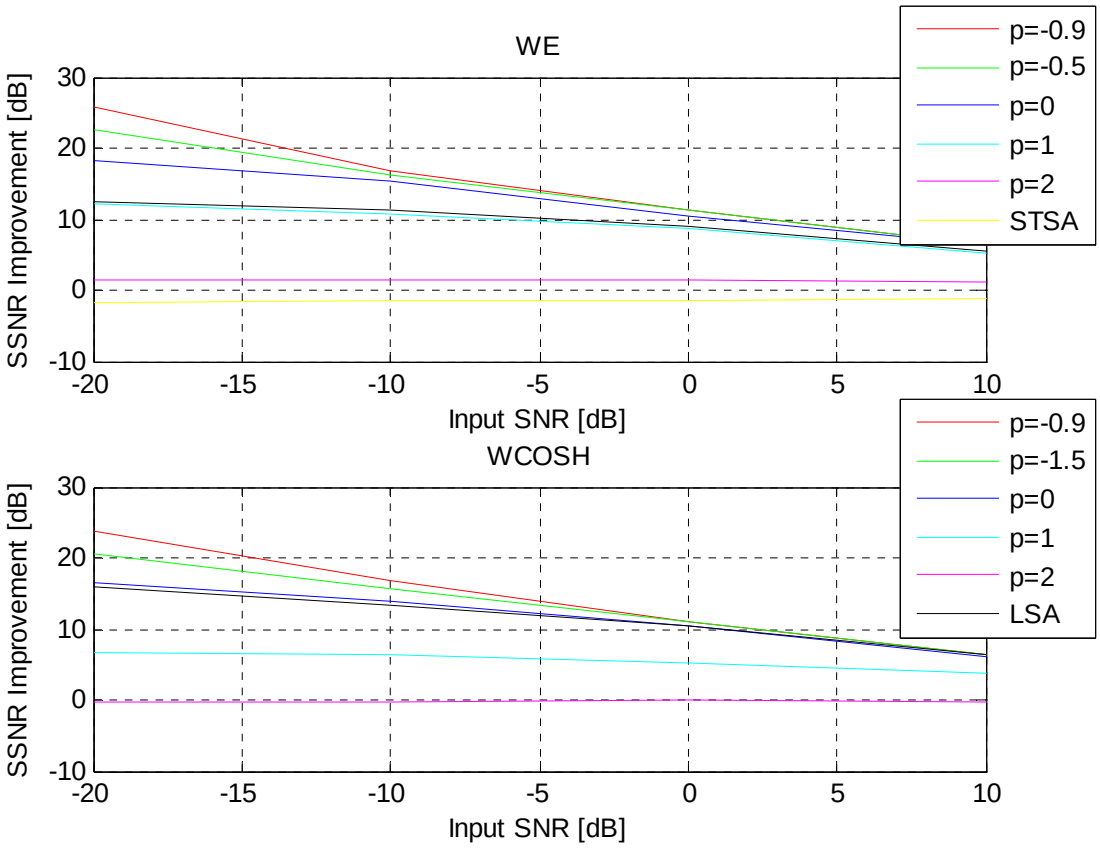

Figure 2-5 SSNR Improvements for Single Channel Weighted Euclidean (WE) and Single Channel Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Single Channel Spectral Phase Estimation

From the enhancement results, the optimal MMSE perceptually-motivated spectral amplitude estimators achieved the best performances with less residual noise and higher speech quality compared to the standard single channel STSA and LSA estimators [12].

2.3.6. Spectral Phase Estimation

In order to reconstruct the clean signal estimate $\hat{s}(t)$, the optimal MMSE spectral phase estimator $\hat{\alpha}$ must be determined from (2.74) that does not alter the optimal MMSE spectral amplitude estimators for $\hat{A}_{STSA}$ and $\hat{A}_{LSA}$ and $\hat{A}_{WE}$ and $\hat{A}_{WCOSH}$ given in (2.86) and (2.118) and (2.137) and (2.138). By formulating the optimal MMSE spectral phase estimator $\hat{\alpha}$ as a constrained Lagrange multiplier problem

$$\begin{aligned} &\min_{e^{j\alpha}} E\left[\left|e^{j\alpha} - e^{j\hat{\alpha}}\right|^2\right] \\ &\text{subject to } \left|e^{j\hat{\alpha}}\right| = 1 \end{aligned} \tag{2.139}$$

or

$$\begin{aligned} &\min_{g,\rho} E\left[\left|e^{j\alpha} - g\right|^2 |Y\right] + \rho\left(|g| - 1\right) \\ &\text{subject to } |g| = 1 \end{aligned} \tag{2.140}$$

with

$$g = e^{j\hat{\alpha}} = g_R + jg_I \tag{2.141}$$

and ρ representing the Lagrange multiplier, the constrained optimal MMSE spectral phase solution is

$$\hat{\alpha} = \tan^{-1}\left(\frac{g_I}{g_R}\right). \tag{2.142}$$

From (2.140), the important relationship between g_R and g_I in (2.141) is

$$\frac{g_I}{g_R} = \frac{E\left[\sin\alpha \middle| Y\right]}{E\left[\cos\alpha \middle| Y\right]}. \tag{2.143}$$

After expanding the terms in the expectation with Euler's identity conditioned on the noisy spectral coefficients Y, (2.140) is written as

$$\begin{aligned} &\min_{g,\rho} E\left[\left|e^{j\alpha} - g\right|^2 \middle| Y\right] + \rho\left(|g| - 1\right) \\ &= \min_{g,\rho} E\left[\left(\cos\alpha - g_R\right)^2 \middle| Y\right] + E\left[\left(\sin\alpha - g_I\right)^2 \middle| Y\right] + \rho\left(g_R^2 + g_I^2\right)^{\frac{1}{2}} - \rho \end{aligned}, \tag{2.144}$$

which requires computation of the partial derivatives of $\frac{\partial\left(E\left[\bullet\right]\right)}{\partial g_R} = 0$ and $\frac{\partial\left(E\left[\bullet\right]\right)}{\partial g_I} = 0$.

The partial derivatives with respect to g_R and g_I are computed to find the solutions of $\frac{\partial\left(E\left[\bullet\right]\right)}{\partial g_R} = 0$ and $\frac{\partial\left(E\left[\bullet\right]\right)}{\partial g_I} = 0$ as

$$g_R\left(2 + \rho\right) = 2E\left[\cos\alpha \middle| Y\right] \tag{2.145}$$

and

$$g_I\left(2 + \rho\right) = 2E\left[\sin\alpha \middle| Y\right]. \tag{2.146}$$

The fundamental relationship between the real and imaginary components is shown in (2.143) with

$$E\left[\cos\alpha \middle| Y\right] = \frac{\int_0^{\infty}\int_0^{2\pi} \cos\alpha\, p\left(Y \middle| A,\alpha\right) p\left(A,\alpha\right) d\alpha dA}{\int_0^{\infty}\int_0^{2\pi} p\left(Y \middle| A,\alpha\right) p\left(A,\alpha\right) d\alpha dA} \tag{2.147}$$

and

$$E\left[\sin\alpha\left|Y\right.\right]=\frac{\int_0^{\infty}\int_0^{2\pi}\sin\alpha\, p\left(Y\left|A,\alpha\right.\right)p\left(A,\alpha\right)d\alpha dA}{\int_0^{\infty}\int_0^{2\pi}p\left(Y\left|A,\alpha\right.\right)p\left(A,\alpha\right)d\alpha dA},\tag{2.148}$$

which closely resemble the integration performed in (2.75), (2.100), (2.127), and (2.128) but with different arguments in the expectation operators. After substituting the statistical models for the speech prior (2.76) and noise likelihood (2.77), (2.147) and (2.148) are rewritten as

$$E\left[\cos\alpha\left|Y\right.\right]=\frac{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\cos\alpha\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha dA}{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha dA}\tag{2.149}$$

and

$$E\left[\sin\alpha\left|Y\right.\right]=\frac{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\sin\alpha\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha dA}{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_x^2}\right)\int_0^{2\pi}\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha dA}.\tag{2.150}$$

By utilizing (2.79), the inner integral over the spectral phase α in (2.149) is expanded as

$$\int_0^{2\pi}\cos\alpha\exp\left(-\frac{\left|Y-Ae^{j\alpha}\right|^2}{\sigma_n^2}\right)d\alpha\propto\int_0^{2\pi}\cos\alpha\exp\left(a\cos\alpha+b\sin\alpha\right)d\alpha.\tag{2.151}$$

Through (2.82), the integral over the spectral phase α in (2.151) is further rewritten as

$$\int_0^{2\pi} \cos\alpha \exp\left(a\cos\alpha + b\sin\alpha\right) d\alpha = \int_0^{2\pi} \cos\alpha \cos\left(\alpha - \psi\right) d\alpha, \qquad (2.152)$$

where

$$\psi = \tan^{-1}\left(\frac{b}{a}\right) \qquad (2.153)$$

and a, b, and $\sqrt{a^2 + b^2}$ are shown in (2.80), (2.81), and (2.83). By using the product-to-sum cosine trigonometric identity, (2.152) simplifies to

$$\begin{aligned}\int_0^{2\pi} \cos\alpha \cos\left(\alpha - \psi\right) d\alpha &= \frac{\sqrt{a^2 + b^2}}{2}\left[\cos\psi \int_0^{2\pi} d\alpha + \int_0^{2\pi} \cos\left(2\alpha - \psi\right) d\alpha\right] \\ &= \pi\sqrt{a^2 + b^2}\cos\psi\end{aligned} \qquad (2.154)$$

since the spectral phase shift of ψ in the second integral over the spectral phase α in (2.154) is irrelevant for the limits of integration. From (2.79) and (2.154), (2.151) is written as

$$\int_0^{2\pi} \cos\alpha \exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|^2}{\sigma_n^2}\right) d\alpha \propto \pi\sqrt{a^2 + b^2}\cos\psi\,. \qquad (2.155)$$

In a similar manner, the inner integral over the spectral phase α in (2.150) is

$$\int_0^{2\pi} \sin\alpha \exp\left(-\frac{\left|Y - Ae^{j\alpha}\right|}{\sigma_n^2}\right) d\alpha \propto \pi\sqrt{a^2 + b^2}\cos\theta\,, \qquad (2.156)$$

where

$$\theta = \sin^{-1}\left(\frac{a}{\sqrt{a^2 + b^2}}\right). \qquad (2.157)$$

Through (2.155) and (2.156), the expectations in (2.149) and (2.150) are written as

$$E\left[\cos\alpha\left|Y\right.\right]=\frac{\sqrt{a^2+b^2}}{2}\cos\psi\frac{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)dA}{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right)dA} \tag{2.158}$$

and

$$E\left[\sin\alpha\left|Y\right.\right]=\frac{\sqrt{a^2+b^2}}{2}\cos\theta\frac{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)dA}{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\frac{Y}{\sigma_n^2}\right|\right)dA} \tag{2.159}$$

with $\frac{1}{\lambda}$ shown in (2.105). By utilizing the expectations from (2.158) and (2.159) and employing the definitions (2.153) and (2.157), the spectral phase estimator is

$$\hat{\alpha}=\tan^{-1}\left(\frac{\cos\theta}{\cos\psi}\right)=\tan^{-1}\left(\frac{b}{a}\right)=\vartheta, \tag{2.160}$$

where a and b are specified in (2.80) and (2.81). Specifically, the single channel optimal MMSE spectral phase estimator is simply the spectral phase of the noisy observation Y.

2.3.7. Complex Real and Imaginary Spectral Component Estimation

In the previous optimal MMSE spectral amplitude estimators, the fundamental assumption has been to model the speech prior and noise likelihood as Gaussian distributions for the spectral amplitude and spectral phase. In contrast, an alternative approach is to determine the MMSE estimate of the real and imaginary spectral

components using Gaussian distributions [13]. By expressing the noisy observation Y in the frequency domain with the real and imaginary components as

$$\begin{aligned} Y &= S + N \\ Y_R + jY_I &= S_R + jS_I + N \end{aligned}, \tag{2.161}$$

the optimal MMSE estimator for the real and imaginary clean spectral components is

$$\begin{aligned} \hat{S}_{R,I} &= E\left[S_{R,I} \middle| Y_{R,I} \right] \\ &= \int_{-\infty}^{\infty} S_{R,I}\, p\left(S_{R,I} \middle| Y_{R,I} \right) dS_{R,I} \\ &= \frac{\int_{-\infty}^{\infty} S_{R,I}\, p\left(Y_{R,I} \middle| S_{R,I} \right) p\left(S_{R,I} \right) dS_{R,I}}{\int_{-\infty}^{\infty} p\left(Y_{R,I} \middle| S_{R,I} \right) p\left(S_{R,I} \right) dS_{R,I}} \end{aligned}, \tag{2.162}$$

where $S_{R,I}$ denotes S_R and S_I but in a more compact form. After substitution of the speech prior

$$p\left(S_{R,I} \right) = \frac{1}{\sqrt{\pi}\sigma_S} \exp\left(-\frac{S_{R,I}^2}{\sigma_S^2} \right) \tag{2.163}$$

and noise likelihood

$$p\left(Y_{R,I} \middle| S_{R,I} \right) = \frac{1}{\sqrt{\pi}\sigma_N} \exp\left(-\frac{\left(Y_{R,I} - S_{R,I} \right)^2}{\sigma_N^2} \right) \tag{2.164}$$

for the Gaussian Noise-Gaussian-Speech statistical models, (2.162) is written as

$$\hat{S}_{R,I} = \frac{\int_{-\infty}^{\infty} S_{R,I} \exp\left(-\left[\frac{\left(Y_{R,I} - S_{R,I}\right)^2}{\sigma_N^2} + \frac{S_{R,I}}{\sigma_S^2} \right] \right) dS_{R,I}}{\int_{-\infty}^{\infty} \exp\left(-\left[\frac{\left(Y_{R,I} - S_{R,I}\right)^2}{\sigma_N^2} + \frac{S_{R,I}}{\sigma_S^2} \right] \right) dS_{R,I}} \tag{2.165}$$

or

$$\hat{S}_{R,I} = \frac{\int_{-\infty}^{\infty} S_{R,I} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I} \left(\frac{Y_{R,I}}{\sigma_N^2} \right) \right) dS_{R,I}}{\int_{-\infty}^{\infty} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I} \left(\frac{Y_{R,I}}{\sigma_N^2} \right) \right) dS_{R,I}}, \tag{2.166}$$

where $\frac{1}{\lambda}$ is defined in (2.105). By splitting the integral in both the numerator and denominator in (2.166) into two separate integrals and utilizing the relationship 3.462.1 in [25],

$$\begin{aligned} &\int_{-\infty}^{\infty} S_{R,I} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I} \left(\frac{Y_{R,I}}{\sigma_N^2} \right) \right) dS_{R,I} \\ &= \left(2\frac{1}{\lambda} \right)^{-1} \exp\left(\frac{\left(\frac{Y_{R,I}}{\sigma_N^2} \right)^2}{2 \cdot \left(\frac{1}{\lambda} \right)} \right) \left[D_{-2}\left(\frac{-\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda} \right)^{\frac{1}{2}}} \right) - D_{-2}\left(\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda} \right)^{\frac{1}{2}}} \right) \right] \end{aligned} \tag{2.167}$$

and

$$\int_{-\infty}^{\infty} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I}\left(\frac{Y_{R,I}}{\sigma_N^2}\right)\right) dS_{R,I}$$
$$= \left(2\frac{1}{\lambda}\right)^{-\frac{1}{2}} \exp\left(\frac{\left(\frac{Y_{R,I}}{\sigma_N^2}\right)^2}{2\cdot\left(\frac{1}{\lambda}\right)}\right)\left[D_{-1}\left(\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right) + D_{-1}\left(\frac{-\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)\right], \quad (2.168)$$

where $D_{\bullet}(\bullet)$ is the parabolic cylinder function defined by 9.240 in [25]. With (2.167) and (2.168), (2.166) is rewritten as

$$\hat{S}_{R,I} = \left(2\frac{1}{\lambda}\right)^{-1/2}\left[\frac{D_{-2}\left(-\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right) - D_{-2}\left(\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)}{D_{-1}\left(\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right) + D_{-1}\left(-\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)}\right], \quad (2.169)$$

where

$$\left(2\frac{1}{\lambda}\right)^{-\frac{1}{2}} = \frac{\sqrt{2}}{2}\left(\frac{\sigma_S^2}{1+\xi}\right)^{\frac{1}{2}}. \quad (2.170)$$

The arguments to the parabolic cylinder function are simplified to

$$\frac{\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{1/2}} = \frac{\frac{\sqrt{\xi}}{\sigma_N}Y_{R,I}}{(1+\xi)^{1/2}} \quad (2.171)$$

and defined as

$$N_{(R,I)\pm} = \sqrt{2}\,\frac{\frac{\sqrt{\xi}}{\sigma_N}Y_{R,I}}{(1+\xi)^{1/2}} \tag{2.172}$$

using the same notation as in [13]. Through the substitution of (2.170) and (2.172), (2.169) is rewritten as

$$\hat{S}_{R,I} = \frac{\sqrt{2}}{2}\left(\frac{\sigma_S^2}{1+\xi}\right)^{\frac{1}{2}}\left[\frac{D_{-2}\left(N_{(R,I)-}\right) - D_{-2}\left(N_{(R,I)+}\right)}{D_{-1}\left(N_{(R,I)+}\right) + D_{-1}\left(N_{(R,I)-}\right)}\right]. \tag{2.173}$$

By simplifying the ratio of the parabolic cylinder functions in (2.173), the ratio is rewritten as

$$\frac{D_{-2}\left(-\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right) - D_{-2}\left(\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)}{D_{-1}\left(\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right) + D_{-1}\left(-\frac{\sqrt{2}\frac{Y_{R,I}}{\sigma_N^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)} = \sqrt{2}\left(\frac{\frac{\sqrt{\xi}}{\sigma_N}Y_{R,I}}{(1+\xi)^{\frac{1}{2}}}\right) = N_{(R,I)+}. \tag{2.174}$$

With (2.174), the estimator for the real and imaginary clean spectral components with Gaussian Noise-Gaussian Speech models in (2.173) is

$$\hat{S}_{(R,I),N-N} = \frac{\xi}{1+\xi}Y_{R,I}. \tag{2.175}$$

In comparison to the Wiener filter and optimal STSA and LSA estimators, the real and imaginary clean spectral component MMSE estimator in (2.175) can generate small improvements in SNR and SSNR performance.

2.4. Dual Channel Enhancement

Besides performing speech enhancement using a single channel microphone, there are many estimation techniques that utilize multiple microphones for improving both the quality and intelligibility. In the subsequent section, dual channel microphone speech enhancement is introduced as an alternative and extension to the traditional and well-established single channel microphone speech enhancement.

2.4.1. Adaptive Noise Cancellation

Adaptive Noise Cancellation (ANC) [2] is an optimal filtering method employed for dual channel microphone signal processing scenarios in either the time domain or frequency domain. Figure 2-6 demonstrates the dual channel ANC method.

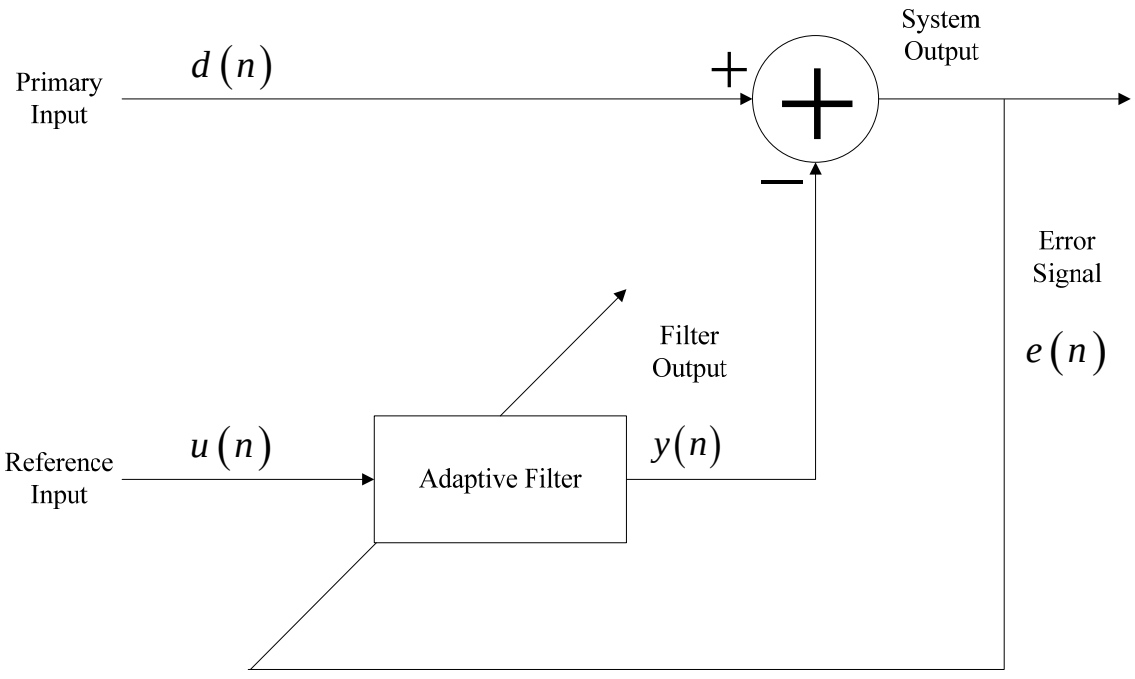

Figure 2-6 Dual Channel Adaptive Noise Cancellation (ANC)

From Figure 2-6, the primary input consists of a desired signal $d(n)$ defined as

$$d(n) = s(n) + v_p(n) \tag{2.176}$$

with the reference input $u(n)$ defined as

$$u(n) = v_r(n), \tag{2.177}$$

where $v_p(n)$ and $v_r(n)$ are correlated noises at the two microphone channels. From (2.177), it is clear that the reference input $u(n)$ contains only noise, not any speech. After the adaptive filter, the output $y(n)$ is

$$y(n) = \sum_{i=0}^{M-1} \hat{w}_i(n) u(n-i) = w^T u \tag{2.178}$$

with weights w that must be determined for each of the M samples of $u(n)$. At the end of the adaptive filtering process, the error $e(n)$ is computed as

$$e(n) = d(n) - y(n) \tag{2.179}$$

and then passed back into the adaptive filter for further processing. The goal is to calculate and update the weights w through numerous iterations to minimize the error in (2.179). Specifically, the minimization of the MSE $E\left[\left|e(n)\right|^2\right]$ will cause the two correlated noises $v_p(n)$ and $v_r(n)$ to match each other and produce ideally zero error with $d(n) = s(n)$. Depending on the algorithm [23], the weight update (e.g., Least Mean Squares or LMS) will typically have the form

$$\hat{w}(n+1) = \hat{w}(n) + \mu u(n) e^{*}(n), \tag{2.180}$$

where μ is the step-size parameter. In the frequency domain, the weight update has a similar form as the time domain weight update in (2.180) [28]. Due to the adaptive capability of the algorithm, speech enhancement systems can process inputs with possibly unknown and non-stationary characteristics and automatically terminate when noticing no further improvement in SNR and SSNR. Overall, the dual channel ANC scheme yields noises and signal distortions that are smaller than single channel optimal filtering configurations.

2.5. Microphone Array

Microphone arrays [3] consist of multiple microphones that require close-spacing of the microphone elements and *a priori* knowledge of the array geometry. Beamforming [8] is a microphone array speech enhancement method that performs spatial filtering to discriminate between the different signals based on the physical location of the sources. The goal is to estimate the signal arriving from a desired look direction in the presence of noise and interfering signals. Beamformers work by forming a scalar output signal as a weighted combination of the source data received at an array of sensors with the weights determining the spatial filtering characteristics. There are two basic classes of beamformers: fixed (conventional) beamformer and adaptive beamformer. While fixed beamformers combine the noisy signals through a time-invariant delay-and-sum or filter-and-sum strategy, adaptive beamformers combine the noisy signals through a time-variant filter-and-sum strategy. By exploiting the spatial dimension of the situation, microphone arrays can acquire a high-quality speech signal without requiring the subject

to talk directly into a single channel microphone [3]. In the next section, fixed beamforming and adaptive beamforming is presented for microphone arrays.

2.5.1. Fixed Beamforming

The simplest microphone array processing strategy for speech enhancement is delay-and-sum beamforming. In order to steer an array of arbitrary configuration and number of microphone sensors M_i for $i = 1,...,M$ to the main lobe of the directivity pattern, the signals received by the array are first delayed to compensate for the path length differences from the source to the various microphone elements and then combined together through a weighting process. Delay-and-sum beamforming can be mathematically expressed in either the time domain as

$$y[n] = \sum_{m=0}^{M-1} w_m x[n - \tau_m] \tag{2.181}$$

or frequency domain as

$$y(f) = \sum_{m=0}^{M-1} w_m x(f) e^{-j2\pi f \tau_m}, \tag{2.182}$$

where w_m is the fixed weight applied to the signal received at a particular microphone m with time delay τ_m. Figure 2-7 shows the basic process of delay-and-sum beamforming.

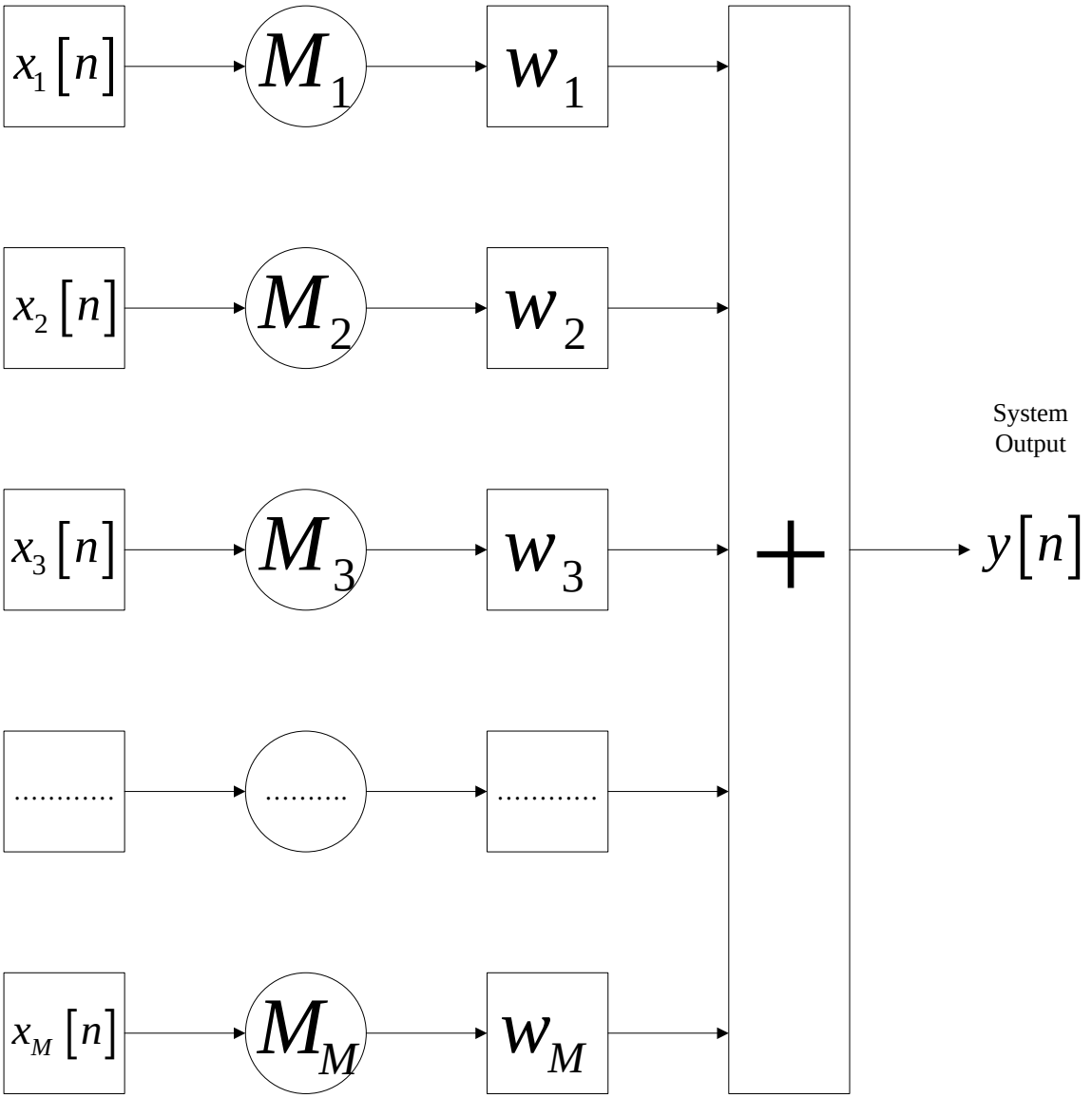

Figure 2-7 Delay-and-Sum Beamforming

While there are many approaches for determining the weights w_m, the most straight-forward approach is to set $w_m = \frac{1}{M}$, where M is the total number of microphones in the array. To estimate the time delay τ_m at each of the microphones, the majority of methods

are based on cross-correlation and similar to the methods used for Time-Difference-of-Arrival (TDOA) with source location [29].

As a generalization of the delay-and-sum beamformer, the filter-and-sum beamformer uses an associated filter for each microphone to filter each of the captured signals before combining them. Filter-and-sum beamforming can be expressed in either the time domain as

$$y[n] = \sum_{m=0}^{M-1} \sum_{p=0}^{P-1} h_m[p] x_m[n - p - \tau_m] \tag{2.183}$$

or frequency domain as

$$y(f) = \sum_{m=0}^{M-1} \sum_{p=0}^{P-1} h_m[p] x_m(f) e^{-j2\pi f(p+\tau_m)}, \tag{2.184}$$

where $h_m[p]$ is the p^{th} tap of the filter associated with the given microphone m. With simply one tap filter $p = 1$ for each microphone m, the time domain and frequency domain filter-and-sum beamformers in (2.183) and (2.184) are equivalent to time domain and frequency domain delay-and-sum beamformers in (2.181) and (2.182).

2.5.2. Adaptive Beamforming

In contrast to fixed beamforming with constant, time invariant weights, adaptive beamforming dynamically adjusts the weights according to some optimization criterion on either a sample-by-sample (time domain) or frame-by-frame (frequency domain) basis. For the optimization criterion, adaptive beamforming techniques usually rely on the minimization of the MSE between the reference signal that is highly correlated to the

desired signal and the output signal. Unfortunately, the normal LMS algorithm often degrades the desired signal since it places no conditions upon the distortion to the desired signal. To deal with the limitation, Frost's algorithm [30] treats the filter estimation process as a problem in a constrained LMS minimization, where the solution still minimizes the MSE while maintaining a specific transfer function for the desired signal. Normally, the constraint is designed to ensure that the response to the desired signal has constant gain and linear phase. As a result, Frost's algorithm belongs to a class of adaptive beamformers known as linearly constrained minimum variance (LCMV) beamformers.

In perhaps the most commonly used LCMV adaptive beamforming technique, the Griffiths-Jim beamformer (GJBF) [31] or generalized sidelobe canceller (GSC) consists of two main processing paths: standard fixed beamformer (FBF) and blocking matrix (BM). The inputs are time-aligned and then passed through a filter-and-sum beamformer to produce the fixed beamformed signal y'_u as

$$y'_u(f) = w_c^T(f)\,x'(f), \tag{2.185}$$

where

$$w(f) = \left[w_1(f),...,w_n(f),...,w_N(f)\right]^T \tag{2.186}$$

are the fixed weights for each of the N microphones with time-aligned input signals

$$x'(f) = \left[x'_1(f),...,x'_n(f),...,x'_N(f)\right]^T. \tag{2.187}$$

To ensure a specified gain and phase response for the desired signal, the output of the FBF is then filtered by a constraint filter h_u with output of the upper path

$$y_u(f) = h_u(f)\, y'_u(f). \tag{2.188}$$

The output is the adaptive section of the beamformer that is driven by outputs from the BM, which removes the desired signal from the lower path. Since the desired signal is common to all of the time-aligned microphone inputs, blocking will occur if the rows of the BM sum to zero. If x'' represents the signals at the output of the BM, then

$$x''(f) = Bx'(f), \tag{2.189}$$

where each row of the blocking matrix sums to zero with linearly independent rows. The standard Griffiths-Jim (GJ) BM B is

$$B = \begin{bmatrix} 1 & -1 & 0 & 0 & . & 0 \\ 0 & 1 & -1 & 0 & . & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & . & 0 & 1 & -1 & 0 \\ 0 & . & 0 & 0 & 1 & -1 \end{bmatrix}, \tag{2.190}$$

where the number of rows in B must be $N-1$ or less because x' can have at most $N-1$ linearly independent components. After application of the BM in (2.190), x'' is adaptively filtered and summed to yield the lower path output y_a as

$$y_a(f) = a^T(f)\, x''(f), \tag{2.191}$$

where a represent the lower path adaptive filters that are updated according to the unconstrained LMS algorithm as

$$a_{k+1}(f) = a_k(f) + \mu y_k(f)\, x''_k(f) \tag{2.192}$$

with step size μ and frame number k. Due to the BM, the lower path output only contains noise signals. Consequently, the overall system output is calculated as the difference of the upper and lower path outputs as

$$y(f) = y_u(f) - y_a(f). \tag{2.193}$$

In practice, adaptive beamformers generally achieve better interference suppression than fixed beamformers [3] but experience a degree of distortion to the desired signal called signal leakage, which is especially problematic for broadband signals such as speech signals since it is difficult to guarantee perfect signal cancellation across a broad frequency range [4].

2.6. Distributed Microphone Enhancement

In the next section, the time domain and frequency domain methods discussed in the previous sections for single channel, dual channel, and microphone array speech enhancement are now generalized for an arbitrary number of microphones dispersed throughout an unknown area for distributed microphone enhancement.

2.6.1. Wiener Filter

In a similar fashion to the optimal single channel MMSE Wiener filter given in [11], the optimal multichannel MMSE Wiener filter has been developed for distributed microphones [10]. For the time domain, the model is

$$\begin{aligned} y_n[k] &= h_n[k] \otimes s[k] + v_n[k] \\ &= x_n[k] + v_n[k] \end{aligned}, \tag{2.194}$$

which consists of N microphone signals $y_n[k]$ of filtered $h_n[k]$ clean speech $s[k]$ and uncorrelated additive noise $v_n[k]$ with $n \in [0,...,N-1]$ at time k. By defining the output signal $z[k]$ as

$$z[k] = \sum_{n=0}^{N-1} w_n^T[k] y_n[k] = w^T[k] y[k], \tag{2.195}$$

the goal is to determine the filter weights $w_n[k]$ for recovery of the clean speech signal $s[k]$ or one of the n filtered clean speech components $x_n[k]$. In (2.195), the time domain quantities $w_n[k]$ and $w[k]$ and $y_n[k]$ and $y[k]$ are defined as

$$w_n[k] = \left[w_n^0[k], w_n^1[k], ..., w_n^{L-1}[k] \right]^T \tag{2.196}$$

and

$$w[k] = \left[w_0^T[k], w_1^T[k], ..., w_{N-1}^T[k] \right]^T \tag{2.197}$$

and

$$y_n[k] = \left[y_n^0[k], y_n^1[k], ..., y_n^{L-1}[k] \right]^T \tag{2.198}$$

and

$$y[k] = \left[y_0^T[k], y_1^T[k], ..., y_{N-1}^T[k] \right]^T. \tag{2.199}$$

Through the definition of the error vector $e = d - z$ with desired response $d = x$, the MSE cost function for optimal filtering is

$$\begin{aligned} J(W) &= E\left[|e|^2\right] = E\left[|d-y|^2\right] \\ &= E\left[d^T d\right] - 2E\left[y^T W d\right] + E\left[y^T W W^T y\right] \end{aligned}. \tag{2.200}$$

By minimizing (2.200) with respect to W, the optimal multichannel Wiener filter is

$$W = R_{yy}^{-1} R_{yd}, \tag{2.201}$$

where $R_{yy} = E\left[yy^T\right]$ and $R_{yd} = E\left[yd^T\right]$ are the $(M \times M)$ correlation matrix of the input signal y and $(M \times M)$ cross-correlation matrix of the input signal y and desired signal d. From (2.198)-(2.199), (2.194) is rewritten as

$$y[k] = x[k] + v[k]. \tag{2.202}$$

Based on the statistical independence of the speech signal $x[k]$ and noise signal $v[k]$

$$R_{xv}[k] = E\left[x[k]v^T[k]\right] = 0, \tag{2.203}$$

the $(M \times M)$ correlation matrix R_{yy} and $(M \times M)$ cross-correlation matrix R_{yd} are expressed as

$$R_{yy}[k] = E\left[y[k]y^T[k]\right] = R_{xx}[k] + R_{vv}[k] \tag{2.204}$$

and

$$R_{yd}[k] = R_{yx}[k] = E\left[y[k]x^T[k]\right] = R_{xx}[k] = R_{yy}[k] - R_{vv}[k]. \tag{2.205}$$

By substitution of (2.204) and (2.205), the optimal multichannel time domain Wiener filter in (2.201) is represented as

$$W = R_{yy}^{-1}[k]\left(R_{yy}[k] - R_{vv}[k]\right), \tag{2.206}$$

where the noise correlation matrix $R_{vv}[k]$ is estimated during speech pauses through a VAD algorithm.

In a similar derivation to the optimal single channel frequency domain Wiener filter, the optimal multichannel frequency domain Wiener filter is calculated from the time domain estimate of the desired signal $Z = X$ as

$$\hat{Z} = W^T Y \tag{2.207}$$

with estimation error

$$\begin{aligned} E &= D - \hat{Z} \\ &= D - W^T Y \end{aligned}, \tag{2.208}$$

where W is the gain function that is applied to all of the noisy observations Y. To compute W, the MSE of (2.208) is defined as

$$\begin{aligned} J &= E\left[|E|^2\right] \\ &= E\left[|D|^2\right] - W^T P_{DY} - W P_{YD} + |W|^2 P_{YY} \end{aligned} \tag{2.209}$$

and minimized with respect to W as

$$\frac{\partial J}{\partial W} = 0 = -P_{YD} + W^* P_{YY}, \tag{2.210}$$

where

$$P_{YD} = E\left[YD^*\right] = P_{XX} \tag{2.211}$$

and

$$P_{DY} = E\left[DY^*\right] = P_{YD} \tag{2.212}$$

and

$$P_{YY} = E\left[YY^*\right] = P_{XX} + P_{NN} \tag{2.213}$$

with power spectrum P_{ij}, where i and j represent the two different signals. By solving (2.210) for the gain function W^* and substituting (2.212) and (2.213), the optimal multichannel frequency domain Wiener filter is written as

$$W^* = \frac{P_{YD}}{P_{YY}} = \frac{P_{XX}}{P_{XX} + P_{NN}} \tag{2.214}$$

or

$$W^* = \frac{\xi}{1+\xi}, \tag{2.215}$$

where ξ is the *a priori* SNR defined as

$$\xi = \frac{P_{XX}}{P_{NN}}. \tag{2.216}$$

The multichannel time domain (2.206) or frequency domain (2.215) Wiener filters have better noise reduction performance than the standard single channel time domain (2.62) or single channel frequency domain (2.70) Wiener filters [10] since they incorporate information from all available channels.

2.6.2. Spectral Amplitude Estimation

As a natural extension to the optimal single channel MMSE STSA given in [5], the optimal multichannel MMSE STSA has been devised for the distributed microphone scenario by Lotter *et al.* [7]. Based on the model

$$\begin{aligned} Y_i &= S_i + N_i \\ R_i e^{j\vartheta_i} &= A_i e^{j\alpha_i} + N_i \end{aligned} \tag{2.217}$$

for $i \in [1,...,M]$ with M microphones and speech prior

$$p(A_i, \alpha_i) = \frac{A_i}{\pi\sigma_{S_i}^2} \exp\left(-\frac{A_i^2}{\sigma_{S_i}^2}\right) \tag{2.218}$$

and noise likelihood

$$p(Y_i \,|A_i, \alpha_i) = \frac{1}{\pi\sigma_{N_i}^2} \exp\left(-\frac{\left|Y_i - A_i e^{j\alpha_i}\right|^2}{\sigma_{N_i}^2}\right) \tag{2.219}$$

statistical models, the multichannel STSA is

$$\begin{aligned} \hat{A}_n^{STSA} &= E\left[A_n \,|Y_1,...,Y_M\right] \\ &= \frac{\int_0^{\infty}\int_0^{2\pi} A_n\, p(Y_1,...,Y_M \,|A_n, \alpha_n)\, p(A_n, \alpha_n)\, d\alpha_n dA_n}{\int_0^{\infty}\int_0^{2\pi} p(Y_1,...,Y_M \,|A_n, \alpha_n)\, p(A_n, \alpha_n)\, d\alpha_n dA_n}. \end{aligned} \tag{2.220}$$

From (2.219), the noisy spectral observations Y_i are independent from each other given a clean spectral amplitude A_n and clean spectral phase α_n at a particular microphone n through the relationship

$$\begin{aligned} p(Y_1,...,Y_M \,|A_n, \alpha_n) &= \prod_{i=1}^{M} p(Y_i \,|A_n, \alpha_n) \\ &= \prod_{i=1}^{M} \frac{1}{\pi\sigma_{N_i}^2} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - \left(\frac{c_i}{c_n}\right) A_n e^{j\alpha_i}\right|^2}{\sigma_{N_i}^2}\right), \end{aligned} \tag{2.221}$$

where $A_i = c_i A$ and $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$ for the true clean spectral amplitude A and true clean spectral variance σ_S^2 with $c_i = 1$ for the attenuation factors. After substitution of (2.218) and (2.221), the result from (2.220) is

$$\hat{A}_n^{STSA} = \frac{\int_0^\infty A_n \exp\left(-\frac{A_n^2}{\sigma_{S_i}^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - \left(\frac{c_i}{c_n}\right) A_n e^{j\alpha_i}\right|^2}{\sigma_{N_i}^2}\right) d\alpha_n dA_n}{\int_0^\infty \exp\left(-\frac{A_n^2}{\sigma_{S_i}^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - \left(\frac{c_i}{c_n}\right) A_n e^{j\alpha_i}\right|^2}{\sigma_{N_i}^2}\right) d\alpha_n dA_n}. \quad (2.222)$$

The integration over the spectral phase α is performed by expansion of the term $\left|Y_i - \left(\frac{c_i}{c_n}\right) A_n e^{j\alpha_i}\right|^2 = \left(Y_i - \left(\frac{c_i}{c_n}\right) A_n e^{j\alpha_i}\right)_R^2 + \left(Y_i - \left(\frac{c_i}{c_n}\right) A_n e^{j\alpha_i}\right)_I^2$ and extracting the constants from the integral as

$$\begin{aligned} &\int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - \left(\frac{c_i}{c_n}\right) A_n e^{j\alpha_i}\right|^2}{\sigma_{N_i}^2}\right) d\alpha_n \\ &= \exp\left(-\sum_{i=1}^{M} \frac{|Y_i|^2 + \left(\frac{c_i}{c_n}\right)^2 A_n^2}{\sigma_{N_i}^2}\right) \int_0^{2\pi} \exp\left(a\cos\alpha + b\sin\alpha\right) d\alpha \end{aligned}, \quad (2.223)$$

where

$$a = \sum_{i=1}^{M} \frac{2c_i A_n}{c_n \sigma_{N_i}^2} \mathrm{Re}\left(Y_i\right) \tag{2.224}$$

and

$$b = \sum_{i=1}^{M} \frac{2c_i A_n}{c_n \sigma_{N_i}^2} \mathrm{Im}\left(Y_i\right). \tag{2.225}$$

From trigonometric identities, the sum of cosine and sine terms with different amplitudes and the same phase is written as in (2.82), where

$$\sqrt{a^2 + b^2} = 2A \left| \sum_{i=1}^{M} \frac{c_i Y_i}{c_n \sigma_{N_i}^2} \right|. \tag{2.226}$$

Since the integral in (2.223) for the spectral phase α is over one full period, the spectral phase shift of $\arctan\left(\frac{b}{a}\right)$ is removed from (2.82). By means of equation 8.431.1 in [25], the integral in (2.223) is rewritten as

$$\int_0^{2\pi} \exp\left(a\cos\alpha + b\sin\alpha\right) d\alpha = 2\pi I_0 \left(2A \left| \sum_{i=1}^{M} \frac{c_i Y_i}{c_n \sigma_{N_i}^2} \right| \right), \tag{2.227}$$

which reduces (2.222)

$$\hat{A}_n^{STSA} = \frac{\int_0^{\infty} A_n^2 \exp\left(-A^2 \frac{1}{\lambda}\right) I_0 \left(2A_n \left| \sum_{i=1}^{M} \frac{c_i Y_i}{c_n \sigma_{N_i}^2} \right| \right) dA_n}{\int_0^{\infty} A_n \exp\left(-A^2 \frac{1}{\lambda}\right) I_0 \left(2A_n \left| \sum_{i=1}^{M} \frac{c_i Y_i}{c_n \sigma_{N_i}^2} \right| \right) dA_n}. \tag{2.228}$$

Through the substitution of equations 8.406.3 and 6.631.1 in [26] and [25], the multichannel STSA is

$$\hat{A}_n^{STSA} = \Gamma(1.5)\sqrt{\frac{\xi_n}{\gamma_n\left(1+\sum_{i=1}^{M}\xi_i\right)}}\,{}_1F_1\left(-0.5;1;-\frac{\left|\sum_{i=1}^{M}\sqrt{\gamma_i\xi_i}e^{j\vartheta_i}\right|^2}{1+\sum_{i=1}^{M}\xi_i}\right)R_n, \quad (2.229)$$

where $A_i = c_i A$ and $\sigma_{S_i}^2 = c_i^2\sigma_S^2$ and ξ_n and γ_n are the *a priori* and *a posteriori* SNRs defined in [7]. In (2.229), it should be noted that the true clean spectral amplitude A_n is estimated at each particular microphone channel n and equivalent to the corresponding single channel spectral amplitude estimator [5]. To completely estimate the clean source signal $\hat{s}(t)$ at each microphone n, the clean spectral phase α is estimated also at each microphone n using the single channel phase estimator $\hat{\alpha}_n = \vartheta_n$ [5]. The multichannel spectral amplitude and single channel spectral phase estimators together provide gains compared to the corresponding single channel estimators [5] from averaging the estimates at each of the microphones n, not by directly estimating the true clean source signal $\hat{s}(t)$ [7].

2.7. Summary

In this chapter, speech enhancement was reviewed for single channel, dual channel, microphone array, and distributed microphone speech enhancement. The goal was to compare and contrast the different algorithms. From the background methods, the theory can be extended now to novel distributed microphone speech enhancement methods.

CHAPTER 3 THEORETICAL METHODS

In this chapter, theoretical methods and derivations are given for distributed microphone statistical estimators for speech enhancement. The basic model is given for the time domain and frequency domain estimators along with explanation about the assumption of the noise field in the surrounding environment. For each of the estimators, the statistical models for both the speech prior and noise likelihood are provided under the assumption of independent noisy observations given the true source signal information along with the final closed-form solution.

3.1. Overview

In a distributed microphone configuration, multiple microphones $i \in [1,...,M]$ capture the attenuated and time delayed coherent clean source signal $c_i s(t-\tau_i)$ corrupted by uncorrelated additive noises $n_i(t)$. By assuming the system can accurately time align the M noisy observations through cross-correlation methods and similar methods used for Time-Difference-of-Arrival (TDOA) with source location [29], the time domain multichannel microphone model is written as

$$y_i(t) = c_i s(t) + n_i(t), \tag{3.1}$$

where $s(t)$ is the true and spatially stationary source signal and $c_i \in [0,1]$ are time-invariant attenuation factors. Figure 3-1 illustrates the basic process of performing speech enhancement on the distributed microphone production model for estimating the true clean source signal $s(t)$.

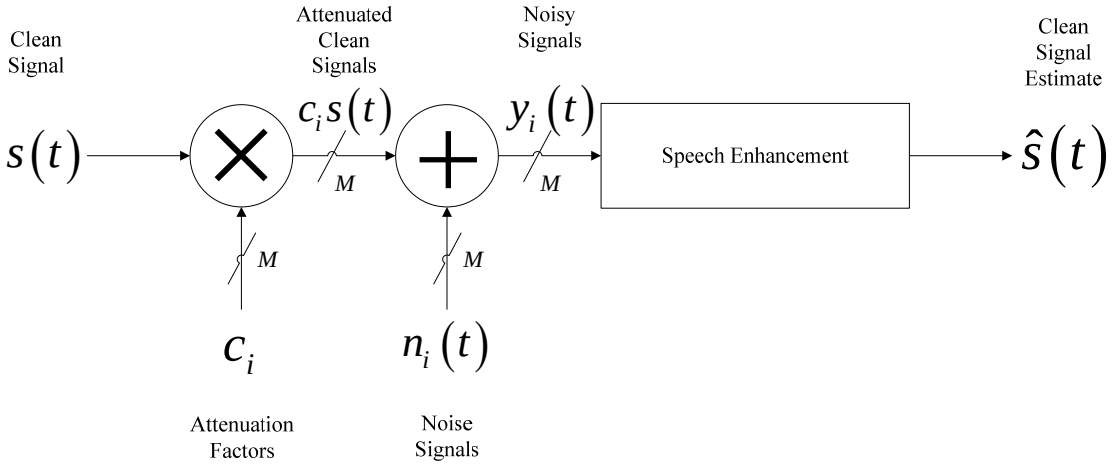

Figure 3-1 Speech Enhancement Applied to Distributed Microphone Production Model

In the frequency domain, (3.1) is expressed in spectral amplitude and spectral phase as

$$\begin{aligned} Y_i(\lambda,k) &= c_i S(\lambda,k) + N_i(\lambda,k) \\ R_i(\lambda,k)e^{j\vartheta_i(\lambda,k)} &= c_i A(\lambda,k)e^{j\alpha(\lambda,k)} + N_i(\lambda,k) \end{aligned} \tag{3.2}$$

or real and imaginary components as

$$\begin{aligned} Y_i(\lambda,k) &= c_i SY_i(\lambda,k) + N_i(\lambda,k) \\ Y_{i,R}(\lambda,k) + jY_{i,I}(\lambda,k) &= c_i\left[S_R(\lambda,k) + jS_I(\lambda,k)\right] + N_i(\lambda,k) \end{aligned}, \tag{3.3}$$

where λ and k represent the frame and frequency bin for each microphone i. To simplify the notation, (3.2) and (3.3) are rewritten without the explicit dependencies as

$$R_i e^{j\vartheta_i} = c_i A e^{j\alpha} + N_i \tag{3.4}$$

or

$$Y_{i,(R,I)} = c_i S_{R,I} + N_i\,. \tag{3.5}$$

For each of the estimators, the fundamental key is to accurately and efficiently estimate the true clean source signal $s(t)$. Figure 3-2 illustrates the operation of speech estimators.

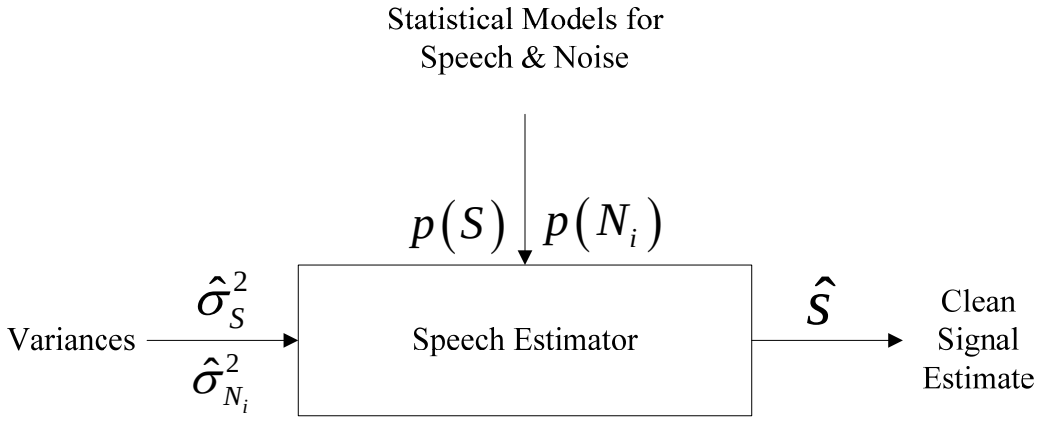

Figure 3-2 Statistical Estimation

Overall, the goal of this work will be to develop the distributed microphone speech enhancement methods shown in Figure 3-1 through the statistical estimation as illustrated in Figure 3-2.

Depending on the noise correlations, there will be more appropriate microphone configurations and speech enhancement methods for a given noisy environment. In general, the majority of large area practical noisy environments (e.g., meeting areas, cafeterias, airport terminals) involve noise situations that are best characterized by a diffuse noise field, where the noise is approximately of equal energy and propagates simultaneously in all directions but has low correlation across the different microphones [4]. The magnitude-squared coherence (MSC) $C_{ij}(f)$ [3] is used to measure the correlation of the various noise signals at any two points in space as a function of

frequency and ratio of power-spectral densities. For a diffuse noise field, the MSC formula is

$$C_{ij}(f) = \frac{\left|P_{ij}(f)\right|^2}{P_{ii}(f)P_{jj}(f)} = \operatorname{sinc}\left(\frac{2\pi f d_{ij}}{c}\right), \tag{3.6}$$

where d_{ij} is the distance between channels i and j and c is the speed of sound. Figure 3-3 illustrates the MSC for a diffuse noise field with different microphone distances.

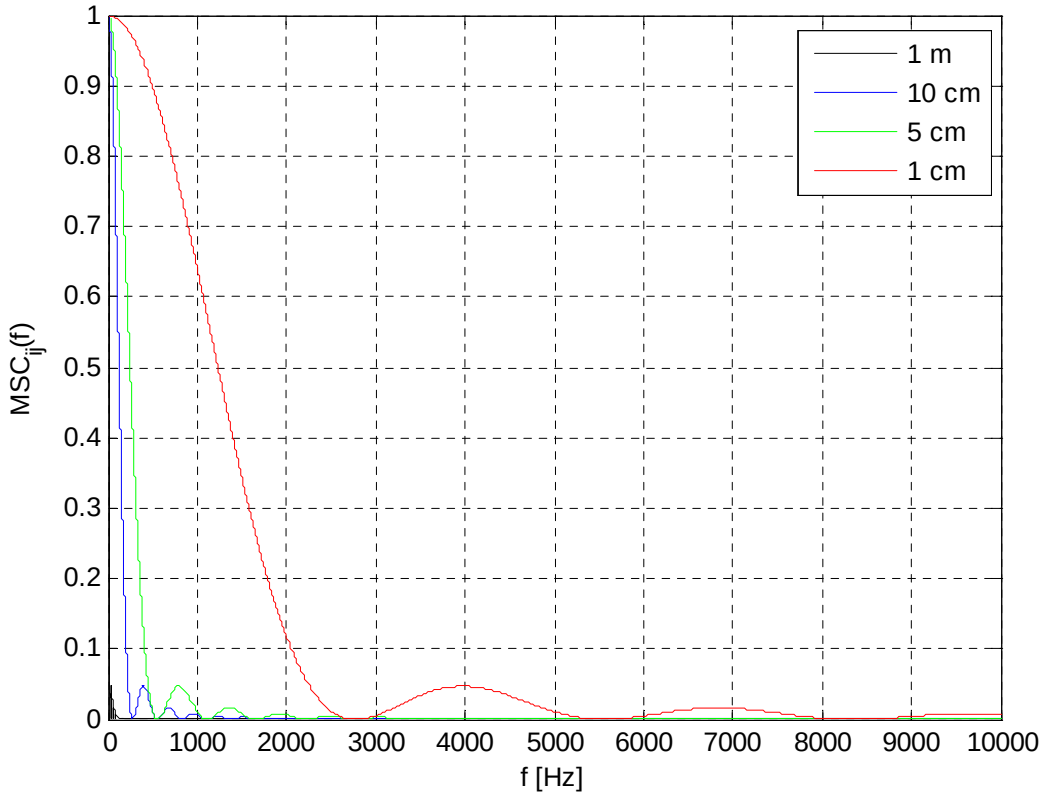

Figure 3-3 Magnitude-Squared Coherence (MSC)

Since the primary energies of speech are mainly concentrated in the 300-3000 Hz frequency range, Figure 3-3 suggests that an assumption of incoherent noise ($C < 0.1$) is justified for microphone spacing above ~14 cm and an assumption of true coherent noise

($C > 0.9$) is justified only for microphone spacing below ~0.4 cm, which is smaller than a typical array. For distributed microphone speech estimators derived in this work, the noise field is assumed to be a diffuse noise field that allows for estimation of the noise statistics at each of the corresponding microphones in the system. The key point to note is that the methods derived in this work will function for an ambient noise field that is even roughly uniform in nature but not quite right for point noise sources, which are correlated and do not have attenuation factor coefficients.

3.2. Time Domain Estimation

For time domain estimation, the MMSE signal estimator in the presence of only white noise will be developed to determine the simple point-by-point estimate of the true source signal s. Figure 3-4 illustrates the distributed microphone time domain speech enhancement estimator for noise reduction through the utilization of all M noisy microphone signals.

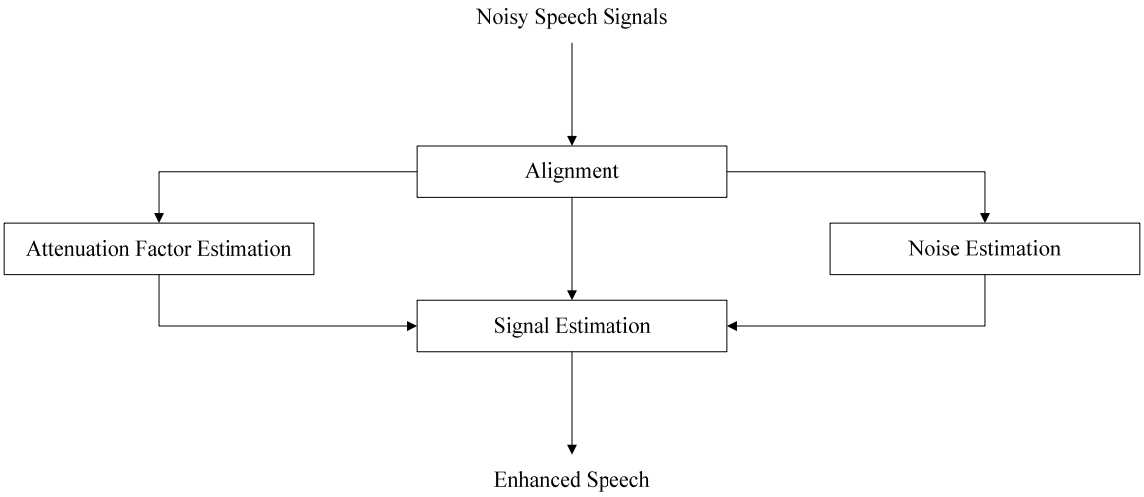

Figure 3-4 Distributed Microphone Time Domain Speech Enhancement System

The goal is to find the MMSE estimate of the true source signal s.

For the statistical models, the noise prior at each microphone i are assumed to be Gaussian models given as

$$p(n_i) = N(0, \sigma_{n_i}), \tag{3.7}$$

where σ_{n_i} is the standard deviation of the noises. From (3.7), the noise likelihood at a particular microphone i is represented as

$$p(y_i | s) = N(c_i s, \sigma_{n_i}), \tag{3.8}$$

where c_i are the attenuation factors for the clean speech s. By assuming a diffuse noise field in a distributed microphone environment, the noise is assumed independent across each of the microphones i, which leads to a product of independent Gaussians for the true source signal s given as

$$\begin{aligned} p(y_1, ..., y_M | s) &= \prod_{i=1}^{M} p(y_i | s) = \prod_{i=1}^{M} N(c_i s, \sigma_{n_i}) \\ &= \prod_{i=1}^{M} \frac{1}{\sigma_{n_i}\sqrt{2\pi}} \exp\left(-\frac{1}{2} \sum_{i=1}^{M} \frac{(y_i - c_i s)^2}{\sigma_{n_i}^2} \right). \end{aligned} \tag{3.9}$$

The conditional joint density in (3.9) can be rewritten as

$$p(y_1, ..., y_M | s) = \prod_{i=1}^{M} \frac{1}{\sigma_{n_i}\sqrt{2\pi}} \exp\left(-\frac{1}{2} \sum_{i=1}^{M} \frac{\left(s - \frac{y_i}{c_i} \right)^2}{\left(\frac{\sigma_{n_i}}{c_i} \right)^2} \right) \tag{3.10}$$

or

$$p\left(y_1,...,y_M\left|s\right.\right)=\prod_{i=1}^{M}\frac{1}{\sigma_{n_i}\sqrt{2\pi}}\exp\left(-\frac{1}{2}\sum_{i=1}^{M}\frac{\left(s-\mu_{s_i}\right)^2}{\sigma_{s_i}^2}\right) \tag{3.11}$$

with mean $\mu_{s_i}=\frac{y_i}{c_i}$ and standard deviation $\sigma_{s_i}=\frac{\sigma_{n_i}}{c_i}$. Based on Bayes theorem

$$p\left(s\left|y_1,...,y_M\right.\right)=\frac{p\left(y_1,...,y_M\left|s\right.\right)p\left(s\right)}{p\left(y_1,...,y_M\right)}, \tag{3.12}$$

the Maximum *A Posteriori* (MAP) and Maximum Likelihood (ML) estimators are determined from the relationships

$$\begin{aligned}\hat{s}_{MAP}&=\arg\max_s p\left(s\left|y_1,...,y_M\right.\right)\\&=\arg\max_s\frac{p\left(y_1,...,y_M\left|s\right.\right)p\left(s\right)}{p\left(y_1,...,y_M\right)}\\&=\arg\max_s p\left(y_1,...,y_M\left|s\right.\right)p\left(s\right)\end{aligned} \tag{3.13}$$

and

$$\hat{s}_{ML}=\arg\max_s p\left(y_1,...,y_M\left|s\right.\right), \tag{3.14}$$

where $p\left(y_1,...,y_M\right)$ is the evidence. By assuming a non-informative prior for $p\left(s\right)$ (i.e., $p\left(s\right)$ is uniform over $\left(-\infty,\infty\right)$), (3.13) can be rewritten as

$$\hat{s}_{MAP}=\arg\max_s p\left(y_1,...,y_M\left|s\right.\right), \tag{3.15}$$

which means $\hat{s}_{MAP}=\hat{s}_{ML}$. With the non-informative prior $p\left(s\right)$ and Gaussian distribution $p\left(y_1,...,y_M\left|s\right.\right)$, the mean (MMSE), mode (MAP), and median (Maximum Absolute Error or MAE) are all equivalent to each other and equal the mean of (3.11) as

$$\hat{s}_{MAP} = \hat{s}_{ML} = \hat{s}_{MMSE} = \mu_s , \tag{3.16}$$

which can be expressed as

$$\begin{aligned} p\left(y_1,...,y_M \,|s\right) &= \prod_{i=1}^{M} \frac{1}{\sigma_{n_i}\sqrt{2\pi}} \exp\left(-\frac{1}{2} \sum_{i=1}^{M} \frac{\left(s-\mu_{s_i}\right)^2}{\sigma_{s_i}^2} \right). \\ &\propto \exp\left(-\frac{1}{2} \frac{\left(s-\mu_s\right)^2}{\sigma_s^2} \right) \end{aligned} \tag{3.17}$$

Through the given statistical models defined in (3.7) and (3.8), the MAP, ML, and MMSE time domain estimators of the true source signal s are given by the conditional mean

$$\hat{s} = \mu_s = E\left[y_1,...,y_M \,|s \right]. \tag{3.18}$$

By completing the square to determine the mean in (3.11) for (3.18), the closed-form solution (see APPENDIX A for details) for $\hat{s}$ is

$$\hat{s} = \mu_s = \sum_{i=1}^{M} w_i y_i , \tag{3.19}$$

where the weights w_i are

$$w_i = \frac{\sigma_s}{\sqrt{\sigma_{y_i}^2 - \sigma_{n_i}^2}} \frac{\prod_{\substack{j=1 \\ j\neq i}}^{M} \sigma_{s_j}^2}{\sum_{i=1}^{M} \prod_{\substack{j=1 \\ j\neq i}}^{M} \sigma_{s_j}^2} \tag{3.20}$$

with the true source standard deviation σ_s and variance $\sigma_{s_i}^2$. In (3.20), the weights w_i are a ratio of SNRs applied to the noisy signals y_i from all other microphones j, except

at that particular microphone $j \neq i$. Ultimately, the noisy observations y_i that contain more clean signal s than noise n will be weighted higher than noisy observations that contain less clean signal than noise.

3.3. Spectral Amplitude Estimation

As opposed to computing the point-by-point estimate of the true source signal $s(t)$ in the time domain and dealing with only white noise, the MMSE estimator will now be developed in the frequency domain for estimation of the true source signal based on the importance of spectral amplitude and spectral phase on quality and intelligibility for any type of noise [20]. Figure 3-5 illustrates the distributed microphone speech enhancement system with the fundamental goal of determining the best estimate of the true source spectral magnitude A and spectral phase α.

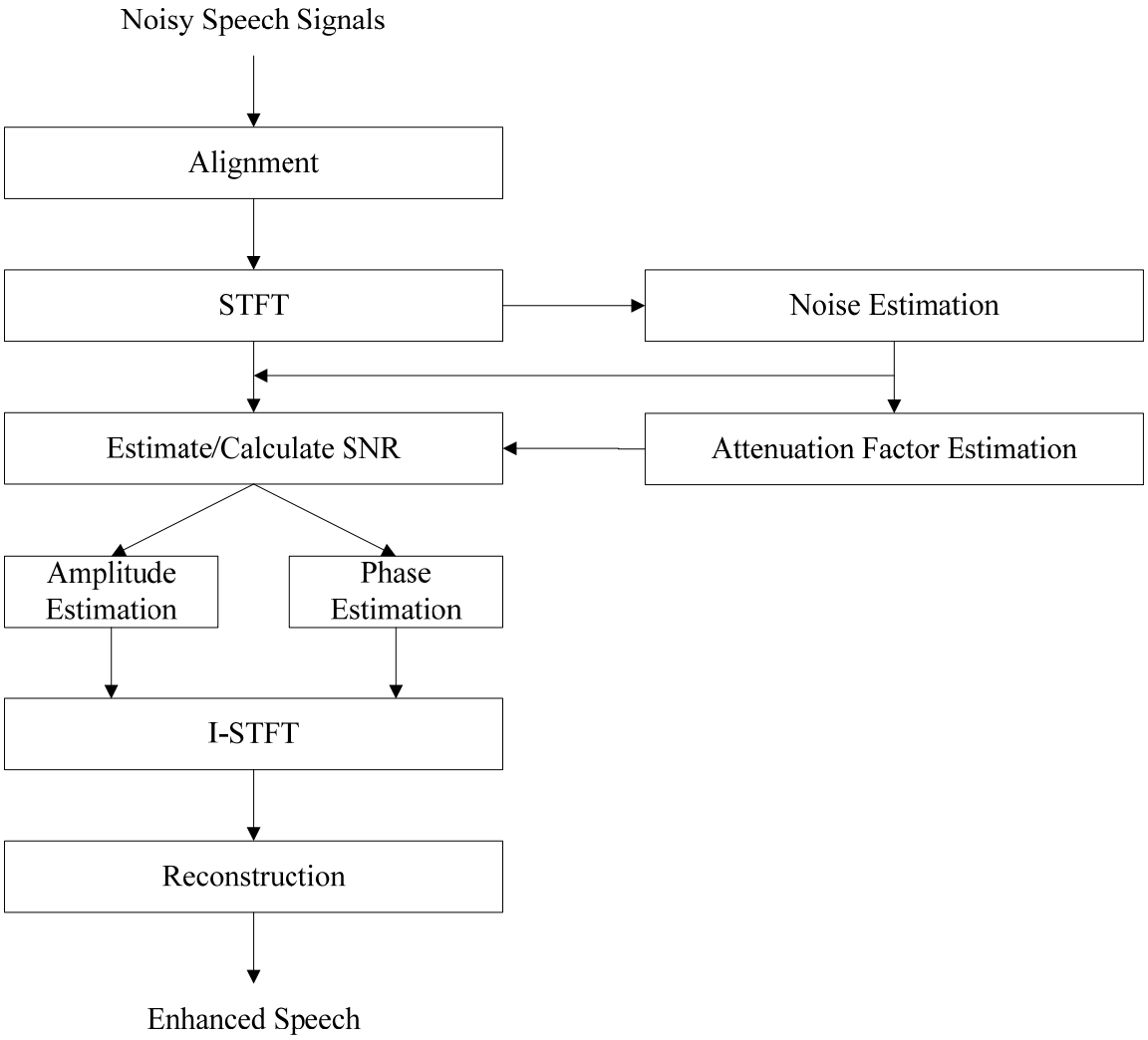

Figure 3-5 Distributed Microphone Spectral Amplitude and Spectral Phase Speech Enhancement System

By assuming Gaussian statistical models, the speech prior and noise likelihood are

$$p(A,\alpha) = \frac{A}{\pi\sigma_S^2}\exp\left(-\frac{A^2}{\sigma_S^2}\right) \tag{3.21}$$

and

$$p\left(Y_i \left| A,\alpha \right.\right) = \frac{1}{\pi\sigma_{N_i}^2}\exp\left(-\frac{\left| Y_i - c_i A e^{j\alpha} \right|^2}{\sigma_{N_i}^2} \right), \tag{3.22}$$

where σ_S^2 and $\sigma_{N_i}^2$ are the speech and noise spectral variances. Based on the assumption of a diffuse noise field for the surrounding environment, the noises are independent at each of the microphone channels, which results in the conditional joint distribution of the noisy spectral observations $\left\{Y_1,...,Y_M\right\}$ written as a product of the independent noisy spectral observations given by

$$\begin{aligned} p\left(Y_1,...,Y_M \left| A,\alpha \right.\right) &= \prod_{i=1}^{M} p\left(Y_i \left| A,\alpha \right.\right) \\ &= \prod_{i=1}^{M} \frac{1}{\pi\sigma_{N_i}^2}\exp\left(-\sum_{i=1}^{M} \frac{\left| Y_i - c_i A e^{j\alpha} \right|^2}{\sigma_{N_i}^2} \right). \end{aligned} \tag{3.23}$$

3.3.1. Short-Time Spectral Amplitude Estimator

Under the given statistical models and following the same method as [7], the MMSE estimate of the STSA is

$$\hat{A}_{STSA} = E\left[A \left| Y_1,...,Y_M \right. \right] = \frac{\int_0^{\infty}\int_0^{2\infty} A p\left(Y_1,...,Y_M \left| A,\alpha \right.\right) p\left(A,\alpha\right) d\alpha dA}{\int_0^{\infty}\int_0^{2\infty} p\left(Y_1,...,Y_M \left| A,\alpha \right.\right) p\left(A,\alpha\right) d\alpha dA}. \tag{3.24}$$

Through substitution of the statistical models in (3.21) and (3.23) into (3.24), the closed-form solution (see APPENDIX B for details) for $\hat{A}_{STSA}$ is

$$\hat{A}_{STSA} = \frac{\Gamma(1.5)}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \frac{{}_1F_1\left(\frac{3}{2};1;v\right)}{{}_1F_1(1;1;v)} \tag{3.25}$$

with

$$\frac{1}{\lambda} = \frac{1}{\sigma_S^2} + \sum_{i=1}^{M} \frac{c_i^2}{\sigma_{N_i}^2} \tag{3.26}$$

and

$$v = \frac{\left| \sum_{i=1}^{M} \frac{c_i Y_i}{\sigma_{N_i}^2} \right|^2}{\frac{1}{\lambda}}, \tag{3.27}$$

where ${}_1F_1(\bullet;\bullet;\bullet)$ denotes the confluent hypergeometric function as described by equation 9.210 in [25]. From the relationship given by equation 9.212.1 in [25], (3.25) is rewritten as

$$\hat{A}_{STSA} = \Gamma(1.5)\left(\frac{\sigma_S^2}{1 + \sum_{i=1}^{M} \xi_i} \right)^{\frac{1}{2}} {}_1F_1\left(-\frac{1}{2};1;-v\right). \tag{3.28}$$

Since the spectral amplitude A and spectral variance σ_S^2 are attenuated at each microphone i by c_i as $A_i = c_i A_i$ and $\sigma_{S_i}^2 = c_i \sigma_S^2$ from the original system model, (3.27) is simplified to

$$v = \frac{\left| \sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} Y_i \right|^2}{1 + \sum_{i=1}^{M} \xi_i}, \tag{3.29}$$

where ξ_i is the *a priori* SNRs and (3.29) is the multichannel extension of v given in [5]. In more specific terms, v is simply a SNR weighted sum of the noisy spectral observations Y_i and normalized by the sum of the *a priori* SNR ξ_i. For a more efficient implementation of the estimator, the confluent hypergeometric function in (3.28) can be replaced by the 0^{th}-Order and 1^{st}-Order modified Bessel functions of the 1^{st} kind given by equations A.1.31a in [32] as

$$\hat{A}_{STSA} = \Gamma(1.5) \left(\frac{\sigma_S^2}{1 + \sum_{i=1}^{M} \xi_i} \right)^{\frac{1}{2}} \exp\left(-\frac{v}{2}\right) \left[(1+v) I_0\left(\frac{v}{2}\right) + v I_1\left(\frac{v}{2}\right) \right]. \tag{3.30}$$

It can be seen that the estimator in (3.30) simplifies to the single channel noise reduction filter [5] for the case of $M = 1$. With rescaling of the attenuation factors to make $c_m = 1$ at a specific reference channel m, (3.30) reduces to the noise reduction filter in [7] for estimating the clean source signal spectral amplitude A_m at each microphone m. The only difference between the method given in [7] by Lotter *el. al.* and the solution given above is that (3.30) is an estimate of the original source spectral amplitude rather than the estimate of the original source spectral amplitude at a particular microphone channel.

3.3.2. Log-Spectral Amplitude Estimator

To obtain a more perceptually relevant criteria function [6], the approach from the previous section is extended to the log-spectral domain as

$$\begin{aligned}\hat{A}_{LSA} &= \exp\left(E\left[\ln(A)\middle|Y_1,...,Y_M\right]\right)\\ &= \exp\left(E\left[Z\middle|Y_1,...,Y_M\right]\right)\end{aligned}, \tag{3.31}$$

where

$$E\left[Z\middle|Y_1,...,Y_M\right] = \frac{d}{d\mu}\left[\Phi_{Z|Y_1,...,Y_M}(\mu)\right]\Big|_{\mu=0} \tag{3.32}$$

and $\Phi_{Z|Y_1,...,Y_M}(\mu) = E\left[A^{\mu}\middle|Y_1,...,Y_M\right]$ is the moment generating function given as

$$\Phi_{Z|Y_1,...,Y_M}(\mu) = \frac{\int_0^{\infty}\int_0^{2\pi} A^{\mu}\, p\left(Y_1,...,Y_M\middle|A,\alpha\right) p(A,\alpha)\, d\alpha\, dA}{\int_0^{\infty}\int_0^{2\pi} p\left(Y_1,...,Y_M\middle|A,\alpha\right) p(A,\alpha)\, d\alpha\, dA}. \tag{3.33}$$

By substitution of the statistical models in (3.21) and (3.23) into (3.33), the closed-form solution (see APPENDIX C for details) for $\Phi_{Z|Y_1,...,Y_M}(\mu)$ is

$$\Phi_{Z|Y_1,...,Y_M}(\mu) = \frac{\Gamma\left(\frac{\mu}{2}+1\right)}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}}\, {}_1F_1\left(-\frac{\mu}{2};1;-\nu\right). \tag{3.34}$$

To complete the derivation of the estimator, it is necessary to differentiate and then perform exponentiation on (3.34). The derivative of (3.34) with respect to μ is written as

$$E\left[Z\left|Y_1,...,Y_M \right.\right] = \frac{d}{d\mu}\left[\Phi_{Z|Y_1,...,Y_M}\left(\mu\right)\right]\Bigg|_{\mu=0}, \tag{3.35}$$

which results in three derivative terms for evaluation at $\mu = 0$. After exponentiation of the three derivative terms (see APPENDIX C for details), the closed-form solution of $\hat{A}_{LSA}$ is written as

$$\hat{A}_{LSA} = \left(\frac{\nu}{1}\right)^{\frac{1}{2}} \exp\left(\frac{1}{2}\int_{\nu}^{\infty}\frac{e^{-t}}{t}dt\right). \tag{3.36}$$

Since the spectral amplitude A and spectral variance σ_S^2 are attenuated at each microphone i by c_i as $A_i = c_i A$ and $\sigma_{S_i}^2 = c_i\sigma_S^2$ from the original system model, the first term in (3.36) can be rewritten as

$$\left(\frac{\nu}{1}\right)^{\frac{1}{2}} = \left(\frac{\sum_{i=1}^{M}\frac{\xi_i}{\gamma_i}}{\sum_{i=1}^{M}\frac{c_i^2}{R_i^2}}\right)^{\frac{1}{2}}\left(\frac{\left|\sum_{i=1}^{M}\frac{\sqrt{\xi_i}}{\sigma_{N_i}}Y_i\right|}{1+\sum_{i=1}^{M}\xi_i}\right). \tag{3.37}$$

The final closed-form solution of the estimator for the source spectral amplitude is

$$\hat{A}_{LSA} = \left(\frac{\sum_{i=1}^{M}\frac{\xi_i}{\gamma_i}}{\sum_{i=1}^{M}\frac{c_i^2}{R_i^2}}\right)^{\frac{1}{2}}\left(\frac{\left|\sum_{i=1}^{M}\frac{\sqrt{\xi_i}}{\sigma_{N_i}}Y_i\right|}{1+\sum_{i=1}^{M}\xi_i}\right)\exp\left(\frac{1}{2}\int_{\nu}^{\nu}\frac{e^{-t}}{t}dt\right), \tag{3.38}$$

where ν is defined as in (3.29) as a weighted and normalized SNR sum of the noisy spectral observations Y_i. It can be seen that the MMSE LSA in (3.38) simplifies to the

single channel MMSE LSA estimator [6] for the case of $M = 1$. As with the estimate of the STSA given in (3.30), (3.38) weights lower the noisy spectral observations Y_i that contain more noise than speech but weights higher the noisy spectral observations that contain more speech than noise to determine an estimate of the LSA.

3.4. Perceptually-Motivated Spectral Amplitude Estimation

By modifying the cost function $d\left(A, \hat{A}\right)$ in (2.120), there are many alternative estimators related to the common STSA and LSA estimator methods of (2.121) and (2.122) for estimating the spectral amplitude of the clean source signal. Since the LSA cost function (2.122) deals with a more perceptual relevant criterion and has produced higher SNR and SSNR improvements in speech quality than the STSA cost function (2.121) for single channel speech enhancement [6], it would seem reasonable that other more perceptually-motivated estimators might also give improved performance for distributed microphone speech enhancement. In single channel enhancement, the best results have occurred with the WE cost function [12]

$$d_{WE}\left(A, \hat{A}\right) = \left(A - \hat{A}\right)^2 A^p \tag{3.39}$$

and WCOSH cost function [12]

$$\begin{aligned} d_{WCOSH}\left(A, \hat{A}\right) &= \left[\frac{1}{2}\left(\frac{A}{\hat{A}} + \frac{\hat{A}}{A}\right) - 1\right] A^p \\ &= \left[\cosh\left(\ln\left(\frac{A}{\hat{A}}\right)\right) - 1\right] A^p \\ &= \left[\cosh\left(\ln\left(A\right) - \ln\left(\hat{A}\right)\right) - 1\right] A^p \end{aligned} \quad . \tag{3.40}$$

By using these cost functions and the basic estimation formulation (3.39) and (3.40), the subsequent true source spectral amplitude estimators for distributed microphone speech enhancement are

$$\hat{A}_{WE} = \frac{\int_0^{\infty}\int_0^{2\pi} A^{p+1} p\left(Y_1,...,Y_M \left| A,\alpha\right.\right) p\left(A,\alpha\right) d\alpha dA}{\int_0^{\infty}\int_0^{2\pi} A^{p} p\left(Y_1,...,Y_M \left| A,\alpha\right.\right) p\left(A,\alpha\right) d\alpha dA} \tag{3.41}$$

and

$$\hat{A}^2_{WCOSH} = \frac{\int_0^{\infty}\int_0^{2\pi} A^{p+1} p\left(Y_1,...,Y_M \left| A,\alpha\right.\right) p\left(A,\alpha\right) d\alpha dA}{\int_0^{\infty}\int_0^{2\pi} A^{p-1} p\left(Y_1,...,Y_M \left| A,\alpha\right.\right) p\left(A,\alpha\right) d\alpha dA}, \tag{3.42}$$

which are valid for the parameters $p_{WE} > 2$ and $p_{WCOSH} > -1$.

3.4.1. Weighted Euclidean Cost Function Spectral Amplitude Estimator

From the given Gaussian statistical models (3.21) and (3.23), the closed-form solution of the true spectral amplitude estimator $\hat{A}_{WE}$ using the WE cost function (see APPENDIX D for details) in (3.41) is

$$\hat{A}_{WE} = \frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+1\right)}\left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i}\right)^{\frac{1}{2}} \frac{{}_1F_1\left(-\left(\frac{p+1}{2}\right);1;-z\right)}{{}_1F_1\left(-\frac{p}{2};1;-z\right)}, \tag{3.43}$$

where z is defined exactly as in (3.29) and $\Gamma(\bullet)$ and ${}_1F_1(\bullet;\bullet;\bullet)$ denote the gamma and confluent hypergeometric functions with free parameter $p_{WE} > 2$. As with the previously

derived estimators, (3.43) decays to the single channel perceptually-motivated Bayesian noise reduction filter using the WE cost function for the case of $M = 1$.

3.4.2. Weighted Cosh Cost Function Spectral Amplitude Estimator

Through the Gaussian speech prior (3.21) and noise likelihood (3.23) models, the closed-form solution of the true spectral amplitude estimator $\hat{A}_{WCOSH}$ using the WCOSH cost function (see APPENDIX E for details) in (3.42) is

$$\hat{A}_{WCOSH} = \sqrt{\frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+\frac{1}{2}\right)}\left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i}\right)\frac{{}_1F_1\left(-\left(\frac{p+1}{2}\right);1;-z\right)}{{}_1F_1\left(-\left(\frac{p-1}{2}\right);1;-z\right)}}, \tag{3.44}$$

where z is defined exactly as in (3.29) and $\Gamma(\bullet)$ and ${}_1F_1(\bullet;\bullet;\bullet)$ denote the gamma and confluent hypergeometric functions equivalently to (3.43) with free parameter $p_{WCOSH} > -1$. In the case of $M = 1$, (3.44) is simply the single channel noise reduction filter [12].

3.5. Spectral Phase Estimation

Besides deriving the MMSE spectral amplitude estimators, the MMSE spectral phase estimator must be derived to construct the enhanced signal. As shown for the MMSE single channel spectral phase estimator in [5], the MMSE estimation of the complex exponential estimator $e^{j\hat{\alpha}}$ results in a non-unity modulus, which produces an altered and a non-optimal estimate of the spectral amplitudes. In order to prevent the optimal spectral phase estimator from affecting the optimal spectral amplitude estimates,

the approach taken in this work is the same constrained optimization formulation from [5] given as

$$\min_{e^{j\hat{\alpha}}} E\left[\left|e^{j\alpha} - e^{j\hat{\alpha}}\right|^2\right], \quad \text{subject to } \left|e^{j\hat{\alpha}}\right| = 1 \tag{3.45}$$

where the magnitude of the complex exponential is constrained to have unity modulus. Through the Lagrange Multiplier optimization method, (3.45) is reformulated as

$$\min_{g,\rho} E\left[\left|e^{j\alpha} - g\right|^2 \middle| Y_1, ..., Y_M\right] + \rho\left(|g| - 1\right) \quad \text{subject to } |g| = 1 \tag{3.46}$$

with

$$g = e^{j\hat{\alpha}} = g_R + jg_I \tag{3.47}$$

and ρ serving as the Lagrange multiplier.

3.5.1. Spectral Phase Estimator

Under the formulation in (3.46), the constrained MMSE solution is

$$\hat{\alpha} = \tan^{-1}\left(\frac{g_I}{g_R}\right). \tag{3.48}$$

From (3.46), the key relationship between the real and imaginary components in (3.47) is

$$\frac{g_I}{g_R} = \frac{E\left[\sin\alpha \middle| Y_1, ..., Y_M\right]}{E\left[\cos\alpha \middle| Y_1, ..., Y_M\right]}. \tag{3.49}$$

By solving and simplifying (3.49) with the attenuation spectral amplitude $A_i = c_i A$ and attenuated spectral variance $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$, the final form of the spectral phase estimator in (3.48) (see APPENDIX F for details) is

$$\hat{\alpha} = \tan^{-1}\left(\frac{\sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} \mathrm{Im}(Y_i)}{\sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} \mathrm{Re}(Y_i)} \right), \tag{3.50}$$

which is a weighted sum of the noisy microphone observations Y_i [33]. As with the spectral amplitude estimators, the spectral phase estimator simplifies to the well-known estimator [5] for the case of $M = 1$, the single channel noisy spectral phase.

3.6. Complex Real and Imaginary Spectral Component Estimation

In contrast to MMSE estimation of the spectral amplitude (3.30) or spectral phase (3.50), the alternative approach is the MMSE estimation of the real and imaginary spectral components of the true source spectrum $S_{R,I}$. As an overview, Figure 3-6 illustrates the distributed microphone speech enhancement estimator for noise reduction through the utilization of all M noisy microphone signals.

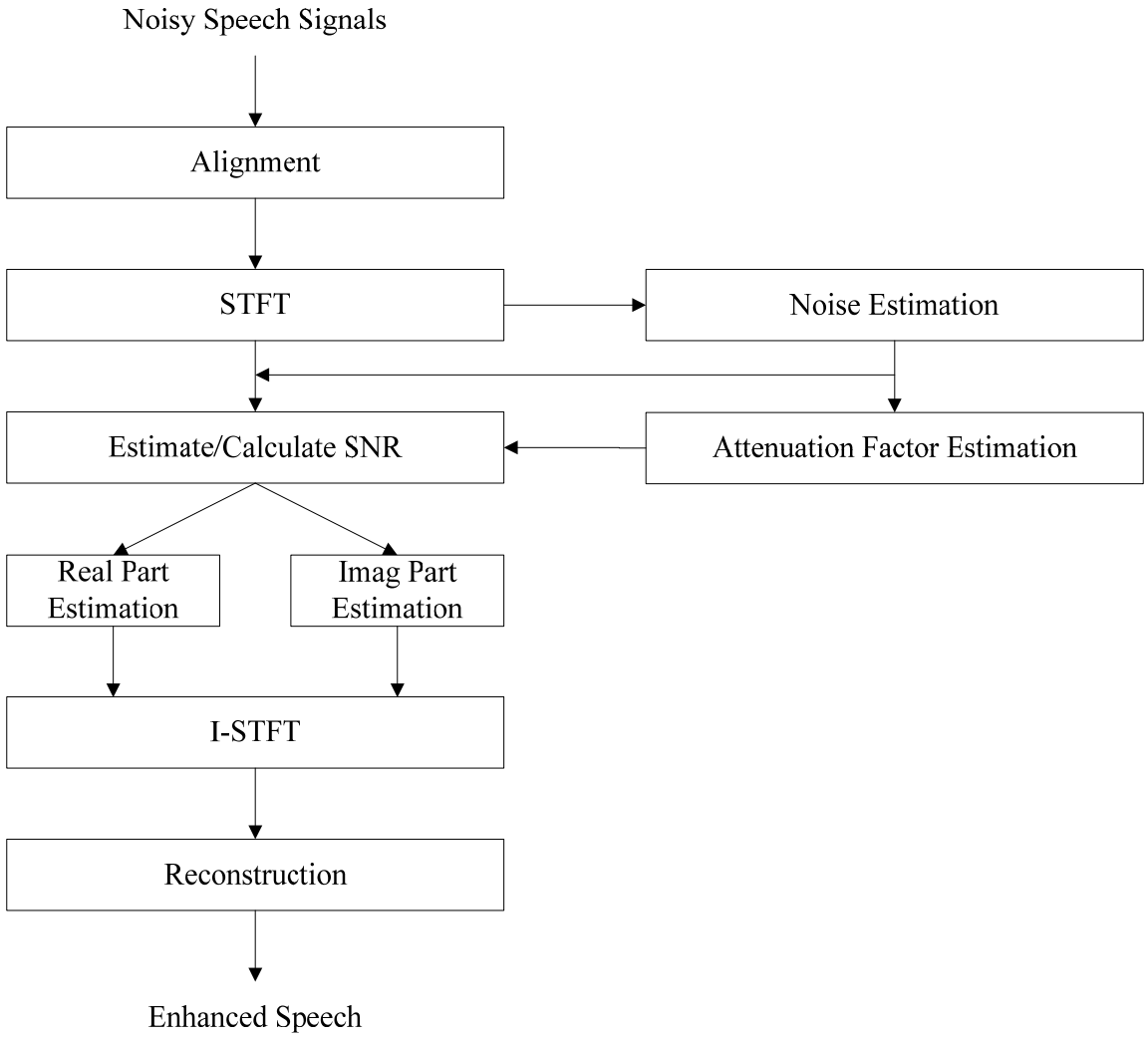

Figure 3-6 Distributed Microphone Complex Real and Imaginary Spectral Component Speech Enhancement System

The goal is to determine the best estimate of the clean source signal real S_R and imaginary S_I spectral components, which are compactly written as $S_{R,I}$.

As with the previous estimators, Gaussian models are assumed for both the speech prior

$$p\left(S_{R,I}\right)=\frac{1}{\sqrt{\pi}\sigma_S}\exp\left(-\frac{S_{R,I}^2}{\sigma_S^2}\right) \tag{3.51}$$

and noise likelihood

$$p\left(Y_{i,(R,I)}\left|S_{R,I}\right.\right)=\frac{1}{\sqrt{\pi}\sigma_{N_i}}\exp\left(-\frac{\left(Y_{i,(R,I)}-c_iS_{R,I}\right)^2}{\sigma_{N_i}^2}\right), \tag{3.52}$$

where σ_S^2 and $\sigma_{N_i}^2$ are the speech and noise spectral variances. Based on the assumption of a diffuse noise field [7] for distributed microphones, the spectral noise components are uncorrelated as

$$\begin{aligned}p\left(Y_{1,(R,I)},...,Y_{M,(R,I)}\left|S_{R,I}\right.\right)&=\prod_{i=1}^{M}p\left(Y_{i,(R,I)}\left|S_{R,I}\right.\right)\\&=\prod_{i=1}^{M}\frac{1}{\sqrt{\pi}\sigma_{N_i}}\exp\left(-\sum_{i=1}^{M}\frac{\left(Y_{i,(R,I)}-c_iS_{R,I}\right)^2}{\sigma_{N_i}^2}\right).\end{aligned} \tag{3.53}$$

Through the statistical models for the speech prior (3.51) and noise likelihood (3.53), the MMSE estimate of the real and imaginary spectral components of the clean spectral source $S_{R,I}$ is

$$\begin{aligned}\hat{S}_{R,I}&=E\left[S_{R,I}\left|Y_{1,(R,I)},...,Y_{M,(R,I)}\right.\right]\\&=\frac{\int_{-\infty}^{\infty}S_{R,I}\,p\left(Y_{1,(R,I)},...,Y_{M,(R,I)}\left|S_{R,I}\right.\right)p\left(S_{R,I}\right)dS_{R,I}}{\int_{-\infty}^{\infty}p\left(Y_{1,(R,I)},...,Y_{M,(R,I)}\left|S_{R,I}\right.\right)p\left(S_{R,I}\right)dS_{R,I}},\end{aligned} \tag{3.54}$$

which simply involves a single integration over the real S_R or imaginary S_I spectral components rather than a double integration over both the clean source signal spectral amplitude A and spectral phase α.

3.6.1. Gaussian Noise-Gaussian Speech Spectral Component Estimator

From the Gaussian statistical models for both the speech prior (3.51) and noise likelihood (3.53), the closed-form solution (see APPENDIX G for details) for $\hat{S}_{R,I}$ is

$$\hat{S}_{R,I} = \frac{\sigma_S \sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} Y_{i,(R,I)}}{1 + \sum_{i=1}^{M} \xi_i}, \tag{3.55}$$

which is applied to the real and imaginary spectral components of the noisy observation signals for distributed microphones. As with the spectral amplitude estimators, (3.55) is simply a weighted SNR sum of the noisy observations $Y_{i,(R,I)}$ and normalized by the sum of the *a priori* SNR ξ_i. For the case of $M = 1$, (3.55) simplifies as

$$\hat{S}_{R,I} = \frac{\xi}{1+\xi} Y_{R,I} \tag{3.56}$$

or

$$\begin{aligned} \hat{S} &= \left|\hat{S}\right| e^{j\angle\hat{S}} \\ &= \hat{S}_R + j\hat{S}_I \\ &= \left(\frac{\xi}{1+\xi}|Y|\right) e^{j\angle Y} \end{aligned}, \tag{3.57}$$

which is the single channel noise reduction Wiener filter as given in [13].

3.7. Summary

In this chapter, the time domain and frequency domain estimators are derived for distributed microphone speech enhancement. For the various spectral amplitude estimators, the spectral phase estimator is a fundamental element in estimating the true clean source signal. In comparison to distributed microphones, the methods derived here

would also apply to directional microphones for an ambient noise field without violation of any of the assumptions; however, the SNR quantities would not be a function of physical distance and would need to employ a different method for estimation.

CHAPTER 4 EXPERIMENTAL WORK

In this chapter, experimental results are presented using the derived statistical estimators for distributed microphone speech enhancement. Overall, there are four basic sets of experiments: enhancement, spectral phase estimation, time alignment, and attenuation factor estimation.

First, in the enhancement experiments, SSNR improvement is measured by increasing the number of microphones with four input SNR levels and three attenuation factor configurations in terms of the various estimators: time domain, spectral amplitude with spectral phase, perceptually-motivated spectral amplitude with spectral phase, and complex real and imaginary spectral component. Fundamentally, the goal is to illustrate that the frequency domain estimators are able to obtain gains in SSNR improvement with an increase in the number of microphones as well as over the simple time domain estimators.

Second, in the spectral phase estimation experiments, SSNR improvement is measured for the STSA and LSA with spectral phase estimators. In essence, the aim is to illustrate the benefit of the newly derived spectral phase estimator over the standard single channel spectral phase estimator.

Third, in the time alignment experiments, the noisy observations are first artificially misaligned by a random number of samples and then processed using the LSA estimator with spectral phase estimator. The purpose is to illustrate the effects on SSNR improvement of operating on unsynchronized frames and then determine whether a

simple cross-correlation technique can compensate for the misalignment and prevent deterioration in the enhancement results.

Fourth, in the attenuation factor experiments, the attenuation factors are given artificial random error and processed using the LSA estimator with spectral phase estimator. The objective is to determine the effects the misestimation of the attenuation factors has on SSNR improvements.

Based on the results, the statistical time domain and frequency domain estimators show SSNR improvements for increasing number of microphones and are robust with respect to time misalignment and attenuation factor, particularly with the inclusion of the spectral phase estimator.

4.1. Overview

The description of the experiments and implementation along with the experimental results are presented for distributed microphone speech enhancement using simulated data.

4.2. Experiments and Implementation

4.2.1. Enhancement

To evaluate the proposed optimal estimators, enhancement experiments were implemented using simulated distributed microphone data. Clean speech was taken from the TIMIT [34] corpus and corrupted by uncorrelated additive stationary zero mean, unity variance white Gaussian noise with input SNR levels range from -20 dB to 10 dB at

increments of 10 dB for 1 to 32 microphones. Table 4-1 shows the three different attenuation factor configurations, where M and i represent the total number of microphones and particular microphone.

Attenuation Factors	**Value**
Unity	$c_i = 1$
Linear	$c_i = \frac{M - i + 1}{M}$
Logarithmic	$c_i = 10^{\left[0, \log_{10}\left(\frac{1}{M}\right) + \frac{(M-2:-1:0)\left(0 - \log_{10}\left(\frac{1}{m}\right)\right)}{floor(M)-1}\right]}$

Table 4-1 Attenuation Factors

To provide an objective measure of the performance, the SSNR metric was used for evaluation, not the SNR metric since it provided very similar overall trends (see APPENDIX H for all supplementary experimental results).

For each of the simulated noisy microphone observations Y_i, the analysis conditions were frames of 256 samples (25.6 ms) with 50% overlap between the corresponding frames using Hanning windows. Noise estimation in either the time domain or frequency domain was performed on an initial silence of 5 frames without any subsequent updating of the time series or spectrum. The DD [5] smoothing approach was utilized to recursively-estimate the *a priori* SNR as

$$\hat{\xi}_i = \frac{\sigma_{S_i}^2}{\sigma_{N_i}^2} = \frac{c_i^2 \sigma_S^2}{\sigma_{N_i}^2} = \alpha_{SNR} \hat{c}_i^2 \frac{\hat{A}^2(\lambda - 1)}{\sigma_{N_i}^2} + (1 - \alpha_{SNR}) P\left[\gamma_i(\lambda) - 1\right], \qquad (4.1)$$

and the *a posteriori* SNR was calculated as

$$\gamma_i = \frac{R_i^2}{\sigma_{N_i}^2} \tag{4.2}$$

for each channel with $\alpha_{SNR} = 0.98$ using thresholds of $\xi_{\min} = 10^{-\frac{25}{10}}$ and $\gamma_{\max} = 40$ (implemented as a floor on $\sigma_{N_i}^2$). By utilizing (4.1) and (4.2) in the four different frequency domain estimators, the spectral amplitude with spectral phase or complex real and imaginary spectral components can be properly estimated for each frame. Based on the estimation, the true source signal estimate $\hat{s}(t)$ was reconstructed through the overlap-add technique. In contrast, the time domain estimator simply involves estimating the noise variance from the initial 5 frames of silence and calculating the noisy variance before determining the weights w_i of the noisy observation y_i. Ultimately, the results are evaluated using the SSNR measure for the enhancement, spectral phase estimation, time alignment error, and attenuation factor estimation experiments.

For the upcoming enhancement experiments, the reference microphone is defined as $m = 1$ with attenuation factor $c_1 = 1$. Given this formulation, Table 4-2, Table 4-3, Table 4-4, Table 4-5, and Table 4-6 show all of the derived estimator equations.

Estimator	**Equation**
Time Domain	$\hat{s}(t) = \sum_{i=1}^{M} \left(\frac{\sqrt{\sigma_{y_1}^2 - \sigma_{n_1}^2}}{\sqrt{\sigma_{y_i}^2 - \sigma_{n_i}^2}} \frac{\prod_{\substack{j=1 \\ j \neq i}}^{M} \sigma_{n_i} \frac{\sqrt{\sigma_{y_1}^2 - \sigma_{n_1}^2}}{\sqrt{\sigma_{y_j}^2 - \sigma_{n_j}^2}}}{\sum_{i=1}^{M} \prod_{\substack{j=1 \\ j \neq i}}^{M} \sigma_{n_i} \frac{\sqrt{\sigma_{y_1}^2 - \sigma_{n_1}^2}}{\sqrt{\sigma_{y_j}^2 - \sigma_{n_j}^2}}} \right) y_i(t)$

Table 4-2 Implementation of the Distributed Microphone Time Domain Estimator

Estimator	Equation
Short-Time Spectral Amplitude (STSA)	$\hat{A}_{STSA} = \Gamma(1.5)\left(\frac{\xi_1}{\gamma_1\left(1+\sum_{i=1}^{M}\xi_i\right)}\right)^{\frac{1}{2}} \exp\left(-\frac{z}{2}\right)\left[(1+z)I_0\left(\frac{z}{2}\right)+zI_1\left(\frac{z}{2}\right)\right]R_1$
Log-Spectral Amplitude (LSA)	$\hat{A}_{LSA} = \left(\frac{\xi_1}{\gamma_1}\right)^{\frac{1}{2}}\left(\frac{\left\lvert\sum_{i=1}^{M}\sqrt{\xi_i\gamma_i}e^{j\vartheta_i}\right\rvert}{1+\sum_{i=1}^{M}\xi_i}\right)\exp\left(\frac{1}{2}\int_{v}^{\infty}\frac{e^{-t}}{t}dt\right)R_1$

Table 4-3 Implementation of the Distributed Microphone Spectral Amplitude Estimators

Estimator	Equation
Weighted Euclidean (WE) Spectral Amplitude	$\hat{A}_{WE} = \frac{1}{\gamma_1}\frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+1\right)}\left(\frac{\xi_1\gamma_1}{1+\sum_{i=1}^{M}\xi_i}\right)^{\frac{1}{2}}\frac{{}_1F_1\left(-\left(\frac{p+1}{2}\right);1;-z\right)}{{}_1F_1\left(-\left(\frac{p-1}{2}\right);1;-z\right)}R_1$
Weighted Cosh (WCOSH) Spectral Amplitude	$\hat{A}_{WCOSH} = \frac{1}{\gamma_1}\sqrt{\frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+\frac{1}{2}\right)}\left(\frac{\xi_1\gamma_1}{1+\sum_{i=1}^{M}\xi_i}\right)\frac{{}_1F_1\left(-\left(\frac{p+1}{2}\right);1;-z\right)}{{}_1F_1\left(-\left(\frac{p-1}{2}\right);1;-z\right)}}R_1$

Table 4-4 Implementation of the Distributed Microphone Perceptually-Motivated Spectral Amplitude Estimators

Estimator	Equation
Spectral Phase	$\hat{\alpha} = \tan^{-1}\left(\frac{\sum_{i=1}^{M}\frac{\sqrt{\xi_i}}{\sigma_{N_i}^2}\mathrm{Im}(Y_i)}{\sum_{i=1}^{M}\frac{\sqrt{\xi_i}}{\sigma_{N_i}^2}\mathrm{Re}(Y_i)}\right)$

Table 4-5 Implementation of the Distributed Microphone Spectral Phase Estimator

Estimator	Equation
Gaussian Noise-Gaussian Speech Spectral Amplitude	$\hat{S}_{R,I} = \frac{\left(\sqrt{\xi_1}\sigma_{N_1}\right)\sum_{i=1}^{M}\frac{\sqrt{\xi_i}}{\sigma_{N_i}}Y_{i,(R,I)}}{1+\sum_{i=1}^{M}\xi_i}$

Table 4-6 Implementation of the Distributed Microphone Complex Real and Imaginary Spectral Component Estimator

For the remaining attenuation factors, c_i is estimated from the signal variances of the noisy observations y_i under the assumed independence of the speech s and noise n_i using

$$\hat{c}_i = \frac{\sqrt{\sigma_{y_i}^2 - \sigma_{n_i}^2}}{\sigma_s} = \frac{\sqrt{\sigma_{y_i}^2 - \sigma_{n_i}^2}}{\sqrt{\sigma_{y_1}^2 - \sigma_{n_1}^2}}, \tag{4.3}$$

which is a relative SNR ratio of the particular microphone i to a reference microphone $m = 1$.

4.2.2. Spectral Phase Estimation

Estimation of the true source spectral phase is a central contribution to the enhancement of the noisy spectral observations Y_i. In order to evaluate the efficacy of the derived spectral phase estimator $\hat{\alpha}$, experiments were run comparing the SSNR using the new spectral phase estimator to the standard single channel spectral phase estimator, which is simply the noisy spectral phase ϑ of the reference channel $m = 1$. The experiment was implemented for a 32 microphone scenario with unity attenuation factors

$c_i = 1$ for the STSA and LSA estimators with spectral phase estimator. SSNR was computed for the enhanced signals.

4.2.3. Time Alignment

For distributed microphones, time alignment of the channels is a significant pre-processing requirement for the estimation of the true source signal $s(t)$. To implement alignment, time delays can be estimated through a variety of methods, which are similar to those methods used for TDOA methods for source localization [29]. The method used here is to select the particular microphone channel with the largest overall signal power as a reference, perform a cross-correlation of the reference against each of the other channels, and use the peak lag of the cross-correlation between the two channels as the time shift for synchronization. Without proper time alignment, the estimators would operate on unsynchronized frames, which would significantly and negatively impact the estimation process.

To evaluate the impact of artificially added misalignment as well as the effectiveness of the selected time alignment method, the noisy spectral observations were artificially time shifted by a random number of samples selected from a zero-mean Gaussian distribution with variance increasing from 0 to 32 with uniform 0.1 increments. The time alignment experiments were implemented for a 32 microphone scenario at 0 dB input SNR with unity attenuation factors $c_i = 1$ for the LSA estimator with spectral phase estimators. SSNR was determined for the enhanced signals.

4.2.4. Attenuation Factor Estimation

To determine an estimate of the true source signal $s(t)$, the attenuation factors c_i must be accurately estimated for calculating the *a priori* SNR ξ_i. Fundamentally, the attenuation factors c_i represent the amplitude reduction between the original acoustic clean source signal s and recorded noisy signals y_i collected at each of the corresponding microphones m. They incorporate several physical and experimental factors such as environmental conditions, distance to the source, directionality and uniformity of the source waveform, and physical relationship between sound pressure level and quantized sample levels. If the source is unidirectional with uniform environment and known air pressure quantization level, then atmospheric models [35] and source localization can be exploited to directly estimate the attenuation factors, which results in an estimate of the true sound pressure level at the source. In most cases, estimation from physical and experimental factors will not be feasible or accurate and the relative attenuation factor ratios between signals can be estimated directly from ratios of noisy signal energies, which leaves only a single degree of freedom. Thus, the value of attenuation factors can be determined by assuming a known c_i at any arbitrary reference microphone.

The impact of artificial error to the attenuation factor on overall enhancement was evaluated by adding random error selected from a zero-mean Gaussian distribution with variance ranging from 0 to 2 in unequal increments to the true constant attenuation factors $c_i = 1$ prior to enhancement. As a flooring mechanism, errors that resulted in

attenuation factors c_i of less than 0 were discarded and randomly re-generated again. The attenuation factor experiments were implemented for a 32 microphone scenario at 0 dB input SNR for the LSA estimator with spectral phase estimator. SSNR was calculated for the enhanced signals.

4.3. Experimental Results

The experimental results for enhancement, spectral phase estimation, time alignment, and attenuation factor estimation are given for the various distributed microphone speech enhancement estimators.

4.3.1. Enhancement

4.3.1.1. Time Domain

Based on an average of 10 trial runs of the same sentence, results of the simulations for unity (Figure 4-1), linear (Figure 4-2), and logarithmic (Figure 4-3) attenuation factors are shown as a function of increasing number of microphones.

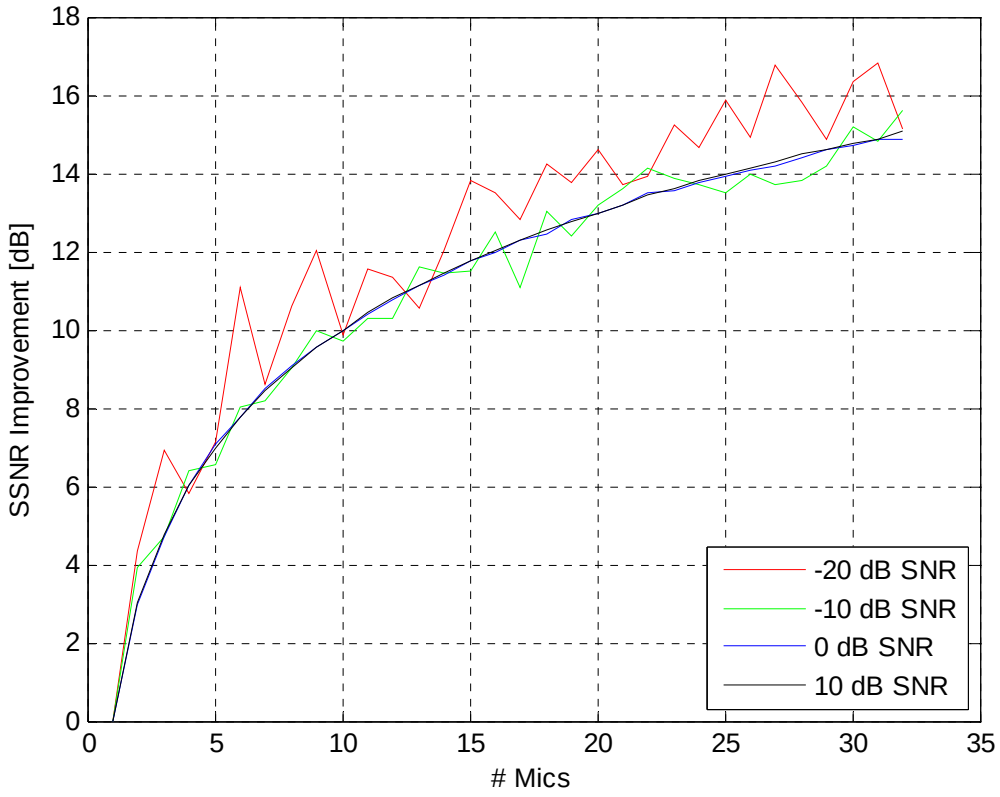

Figure 4-1 SSNR Improvements for Time Domain Estimation (Unity Attenuation Factors)

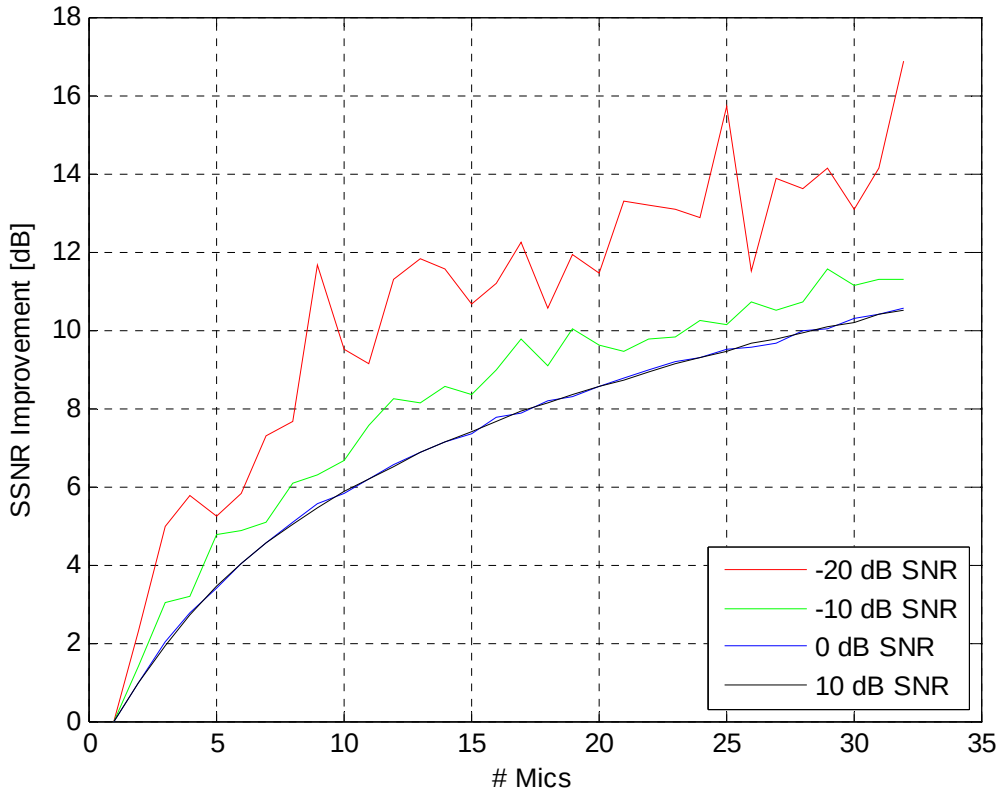

Figure 4-2 SSNR Improvements for Time Domain Estimation (Linear Attenuation Factors)

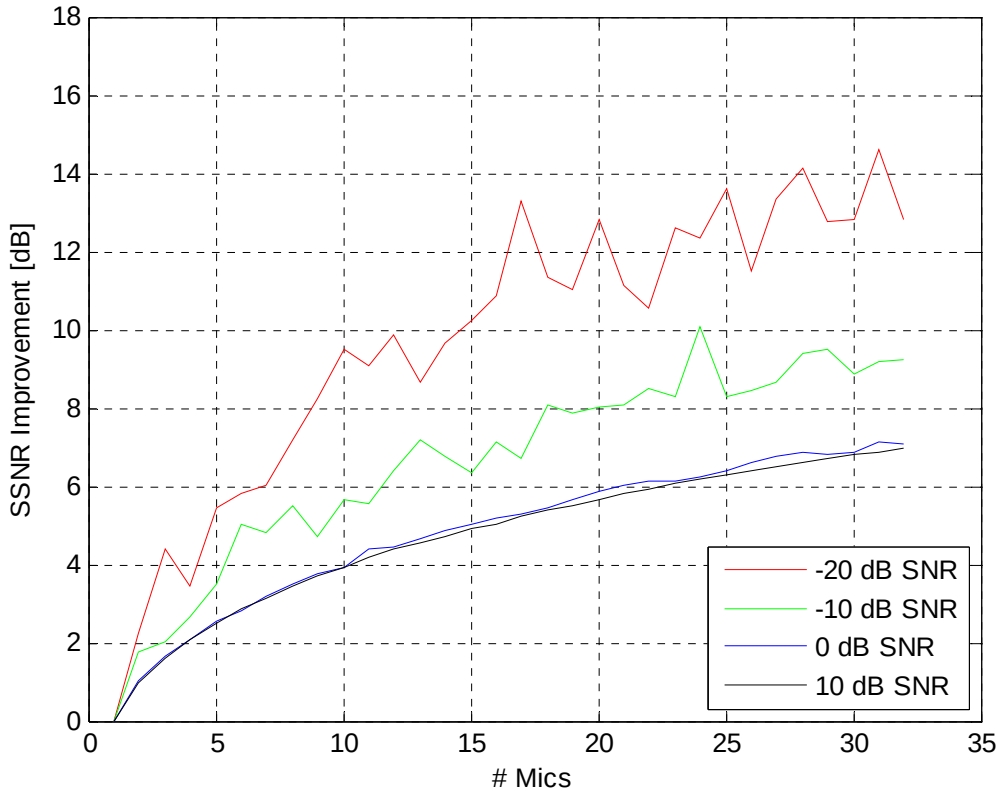

Figure 4-3 SSNR Improvements for Time Domain Estimation (Logarithmic Attenuation Factors)

The unity attenuation factors produced the best overall enhancement from the input SNRs and varying number of microphones. For the lower SNR levels (e.g., -20 dB and -10 dB), there was a great deal of fluctuation in the results across all attenuation factors; however, the higher SNR levels (e.g., 0 dB and 10 dB) showed a consist increase in the output SSNR improvement with increasing number of microphones. At -10 dB, 0 dB, and 10 dB, the unity attenuation factor case exceeded the linear and logarithmic attenuation factor cases by an average of 4-6 dB (-10 dB) and 3-8 dB (0 dB and 10 dB). In Figure 4-4, the time-series and spectrograms are given for the true source signal, noisy signal, and true source signal estimate.

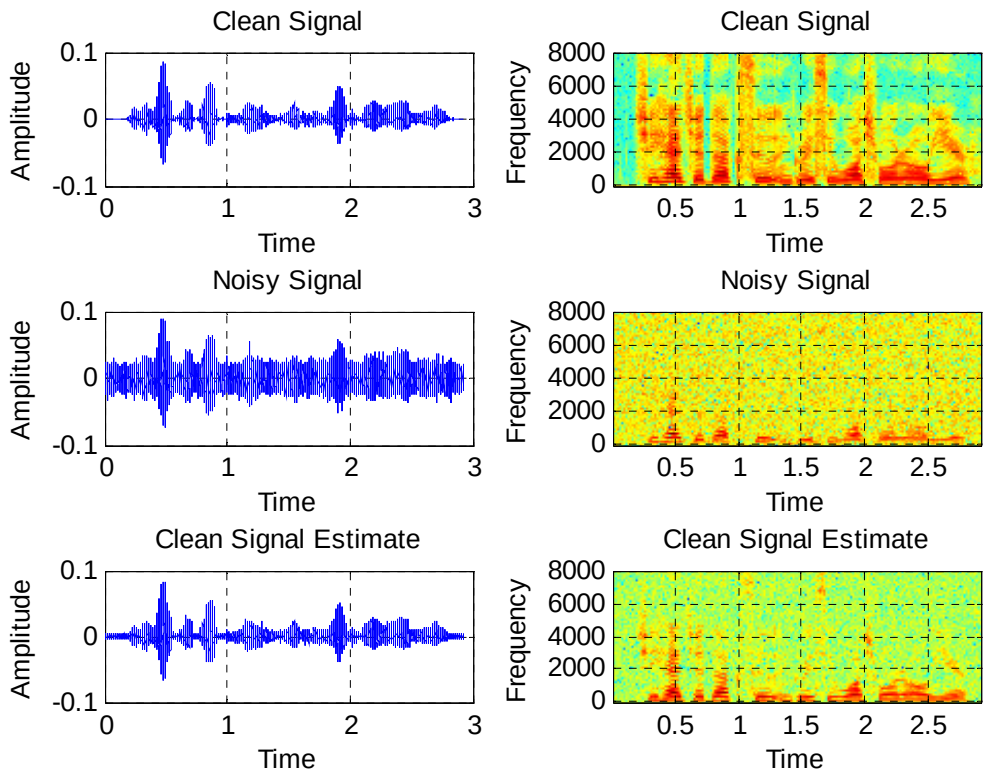

Figure 4-4 Time-Series and Spectrograms of Clean Signal, Noisy Signal, and Clean Signal Estimate for Time Domain Estimation (Unity Attenuation Factors, 0 dB Input SNR, 32 Microphones)

The time domain estimator was able to recover only some of the speech energy in the frequencies ranging from 0 Hz to 4000 Hz, but the higher frequencies are nearly non-existent in the spectrogram of the enhanced signal. Clearly, the time domain estimator only provides for baseline of enhancement. Overall, the time domain estimator is a simple time domain enhancement method that can be compared to the subsequent frequency domain estimators.

4.3.1.2. Spectral Amplitude and Spectral Phase

The results of the simulations for the STSA and LSA estimators with spectral phase estimator were averaged over 10 trial runs of the same sentence and are shown for unity (Figure 4-5), linear (Figure 4-6), and logarithmic (Figure 4-7) attenuation factors as a function of increasing number of microphone channels.

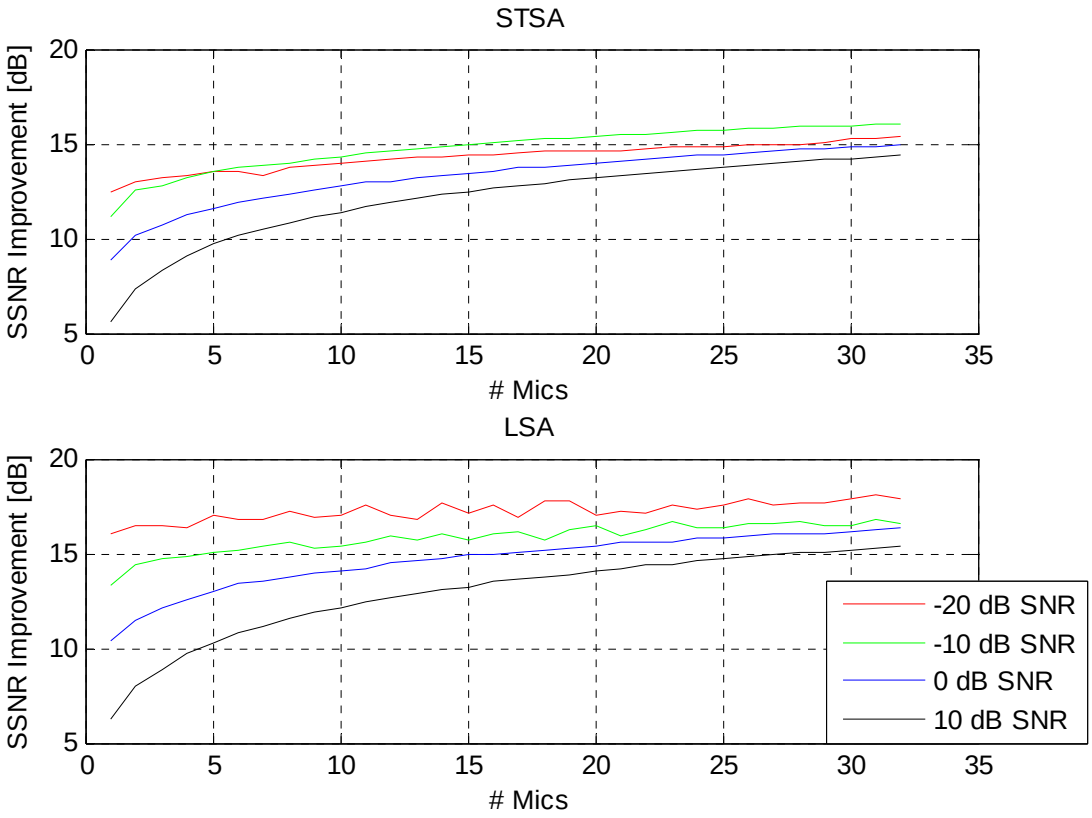

Figure 4-5 SSNR Improvements for Short-Time Spectral Amplitude (STSA) and Log-Spectral Amplitude (LSA) Estimation with Spectral Phase Estimation (Unity Attenuation Factors)

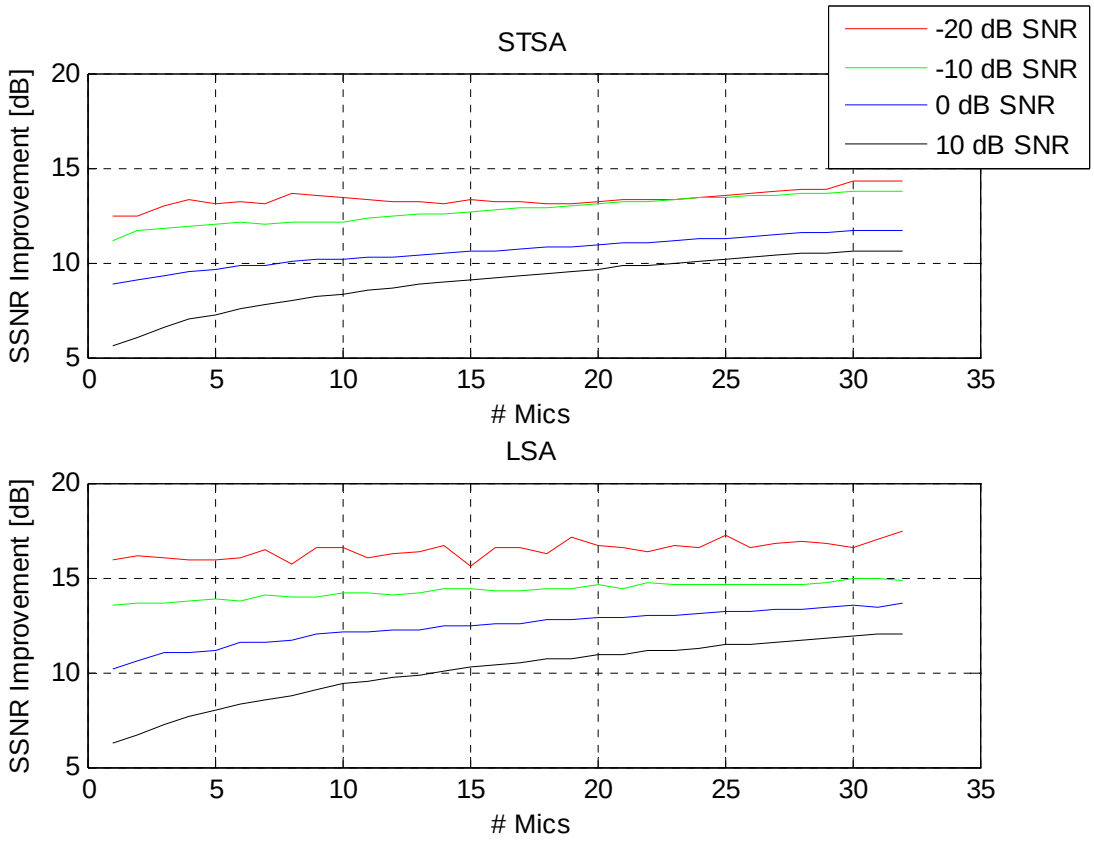

Figure 4-6 SSNR Improvements for Short-Time Spectral Amplitude (STSA) and Log-Spectral Amplitude (LSA) Estimation with Spectral Phase Estimation (Linear Attenuation Factors)

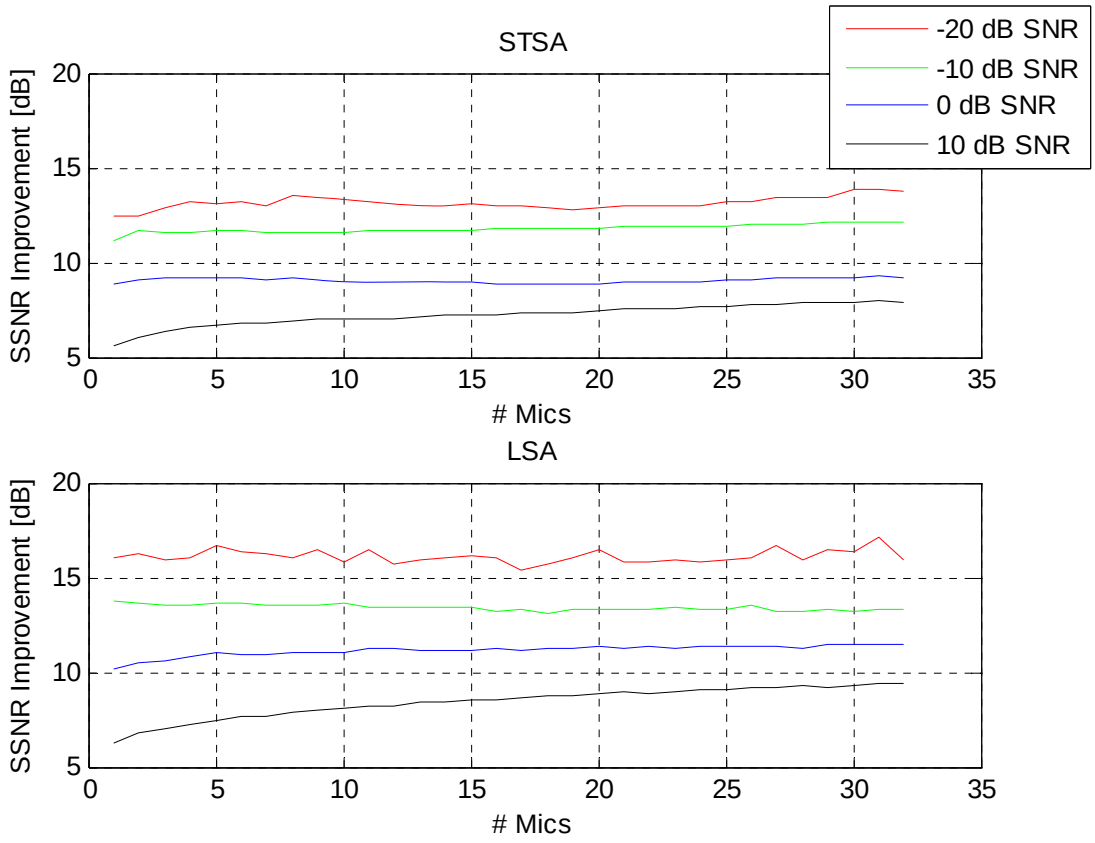

Figure 4-7 SSNR Improvements for Short-Time Spectral Amplitude (STSA) and Log-Spectral Amplitude (LSA) Estimation with Spectral Phase Estimation (Logarithmic Attenuation Factors)

The unity attenuation factor case with LSA estimator with spectral phase estimator provided the best estimation results for all input SNR levels and varying number of microphones. The amount of SSNR improvement due to increasing number of channels was most significant for the less noisy SNR cases of 10 dB and 0 dB with the highest total improvement of 9 dB SSNR increase for 32 microphones versus 1 microphone in the 10 dB SNR and unity attenuation factor case. For the lowest input SNR levels of -10 dB and -20 dB, the improvements were not nearly as pronounced over the baseline single channel enhancement. From the different attenuation factors, the LSA estimator with

spectral phase estimator produced 2 dB SSNR improvement to 18 dB at -20 dB input SNR (unity attenuation factor), 1-3 dB SSNR improvement (linear attenuation factor), and 2 dB SSNR improvement (logarithmic attenuation factor) over the STSA estimator with spectral phase estimator. In comparison to the attenuation factor cases in Figure 4-1, Figure 4-2, and Figure 4-3, the STSA and LSA estimators with spectral phase estimator yielded results that were slightly better than the time domain estimator by approximately 1-3 dB for Gaussian white noise. Figure 4-8 and Figure 4-9 show the time-series and spectrograms for the true source signal, noisy signal, and true source signal estimate.

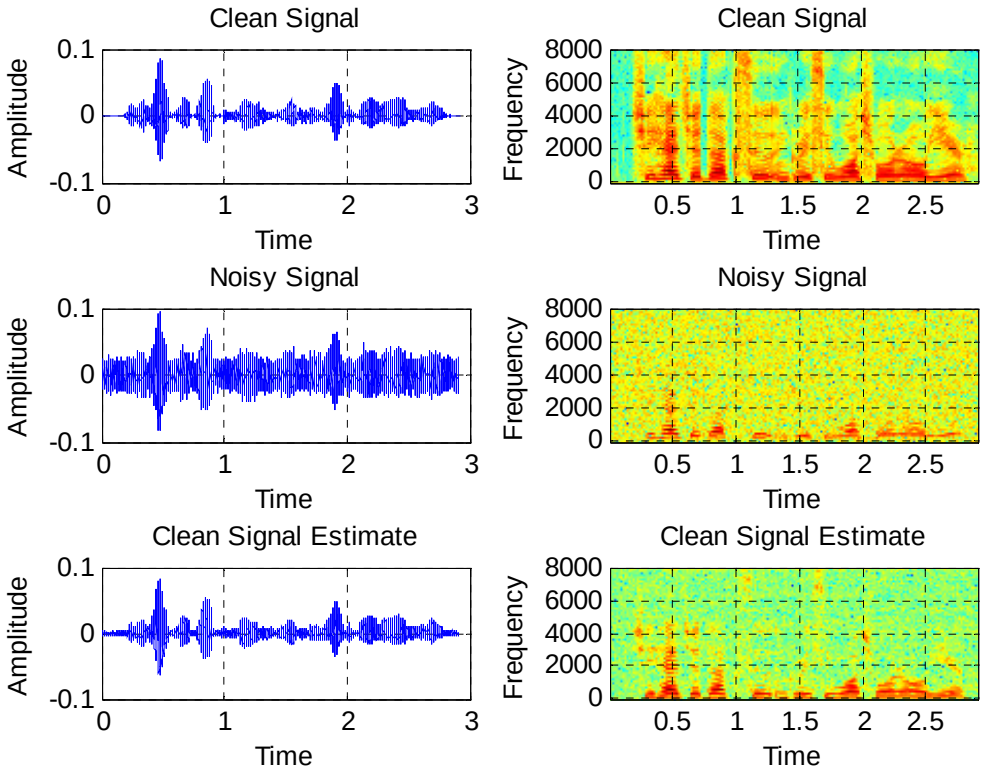

Figure 4-8 Time-Series and Spectrograms of Clean Signal, Noisy Signal, and Clean Signal Estimate for Spectral Amplitude (STSA) Estimation with Spectral Phase Estimation (Unity Attenuation Factors, 0 dB Input SNR, 32 Microphones)

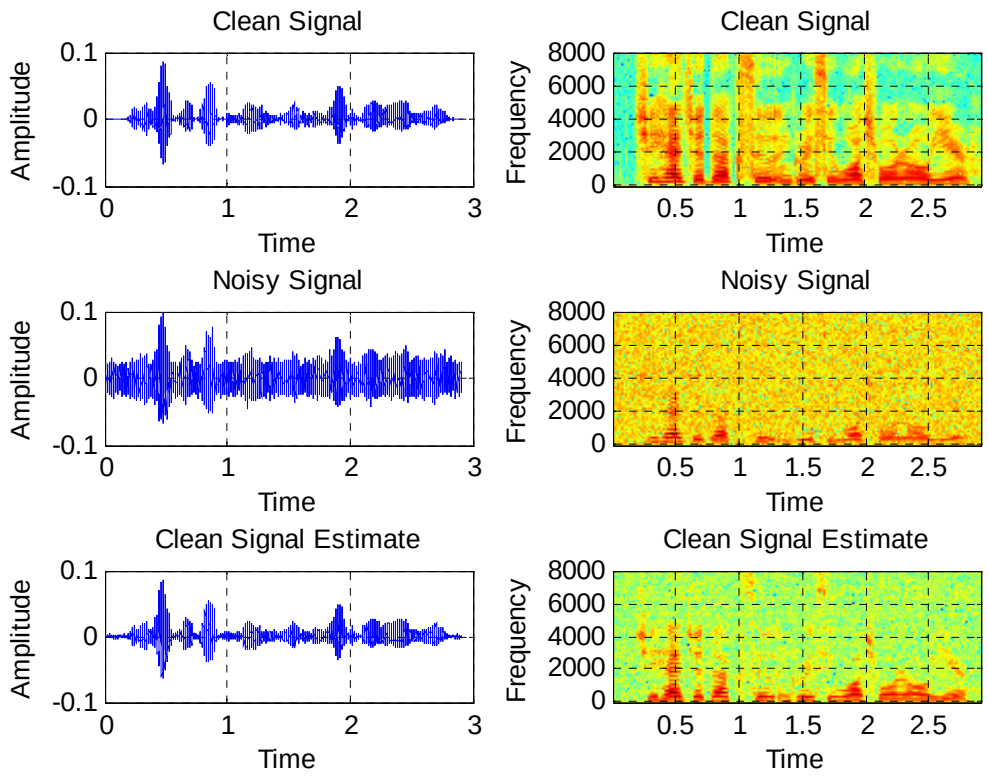

Figure 4-9 Time-Series and Spectrograms of Clean Signal, Noisy Signal, and Clean Signal Estimate for Log-Spectral Amplitude (LSA) Estimation with Spectral Phase Estimation (Unity Attenuation Factors, 0 dB Input SNR, 32 Microphones)

By comparing Figure 4-8 and Figure 4-9, the spectrograms of the enhanced signals from the STSA and LSA estimators with spectral phase estimator contained much more of the speech energies than the spectrogram of the enhanced signal with the time domain estimator in Figure 4-4, particularly at higher frequencies. The reason for the clearer spectrograms from the STSA and LSA estimators with spectral phase estimator is that frequency domain estimators can better capitalize on the important spectrum information of the signal. Whereas the LSA estimator with spectral phase estimator recovered additional clean speech frequencies over the STSA estimator with spectral phase

estimator, the STSA estimator with spectral phase estimator reduced more of the background noise than the LSA estimator with spectral phase estimator. Ultimately, the STSA and LSA estimators with spectral phase estimator are frequency domain enhancement methods that show significant SSNR improvements over the simple time domain enhancement method.

4.3.1.3. Perceptually-Motivated Spectral Amplitude and Spectral Phase

The simulation results for the WE and WCOSH estimators with spectral phase estimator are shown for 1 trial run in Figure 4-10, Figure 4-11, and Figure 4-12 and Figure 4-13, Figure 4-14, and Figure 4-15 as a function of the number of microphones and different attenuation factors.

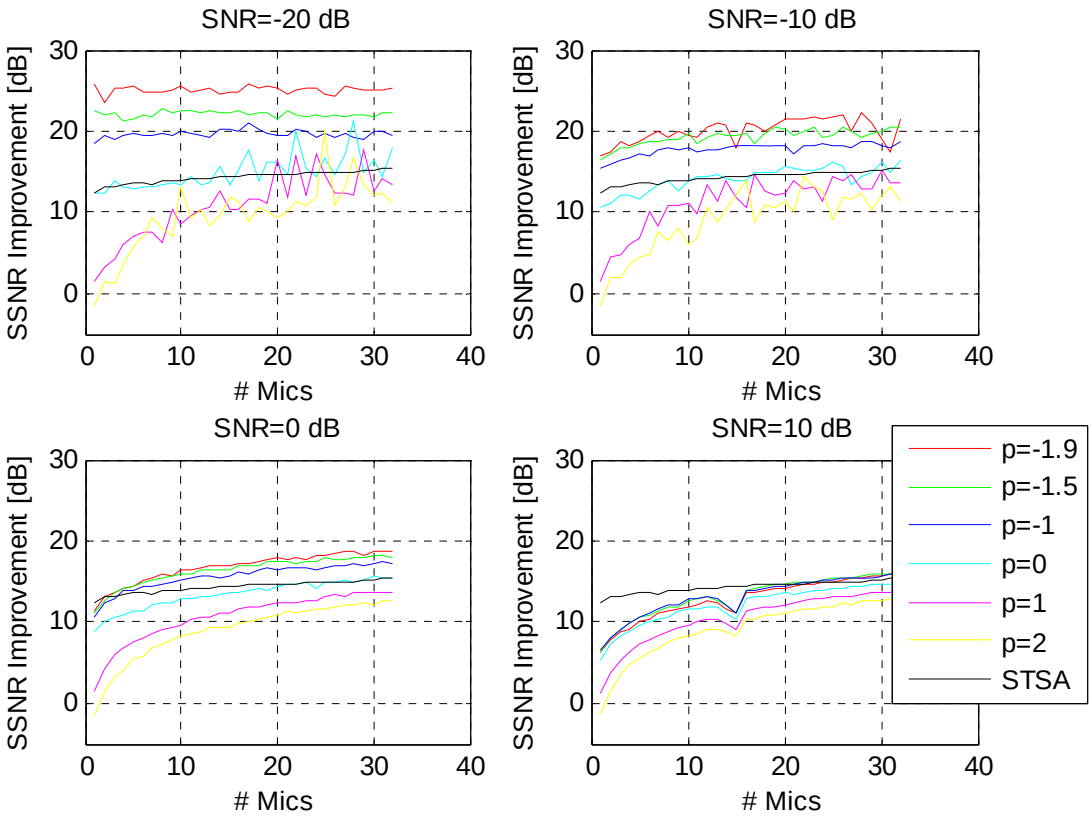

Figure 4-10 SSNR Improvements for Weighted Euclidean (WE) Spectral Amplitude Estimation with Spectral Phase Estimation (Unity Attenuation Factors)

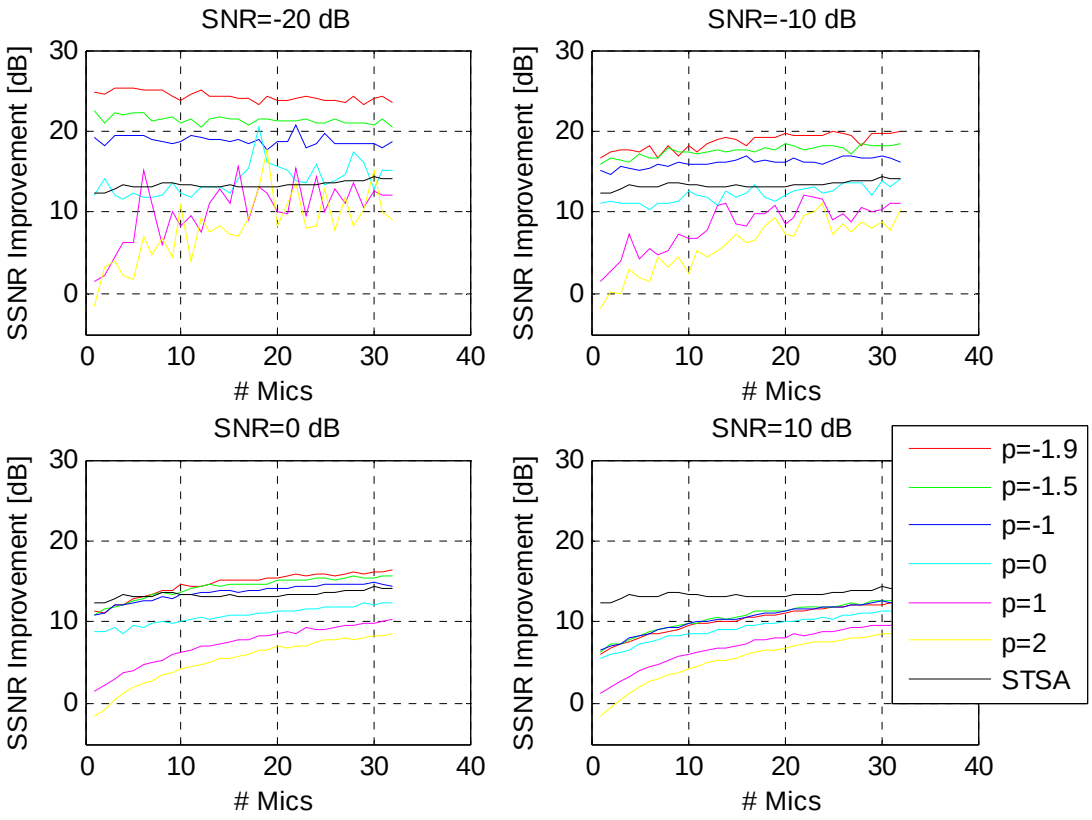

Figure 4-11 SSNR Improvements for Weighted Euclidean (WE) Spectral Amplitude Estimation with Spectral Phase Estimation (Linear Attenuation Factors)

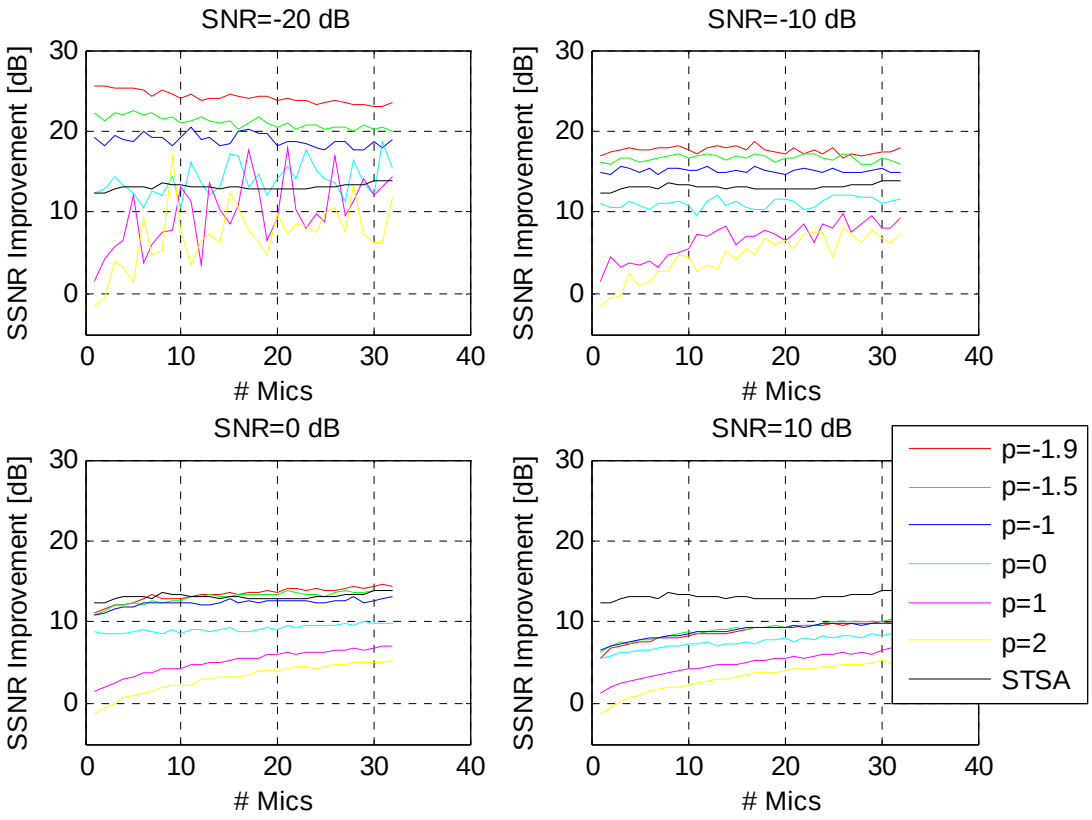

Figure 4-12 SSNR Improvements for Weighted Euclidean (WE) Spectral Amplitude Estimation with Spectral Phase Estimation (Logarithmic Attenuation Factors)

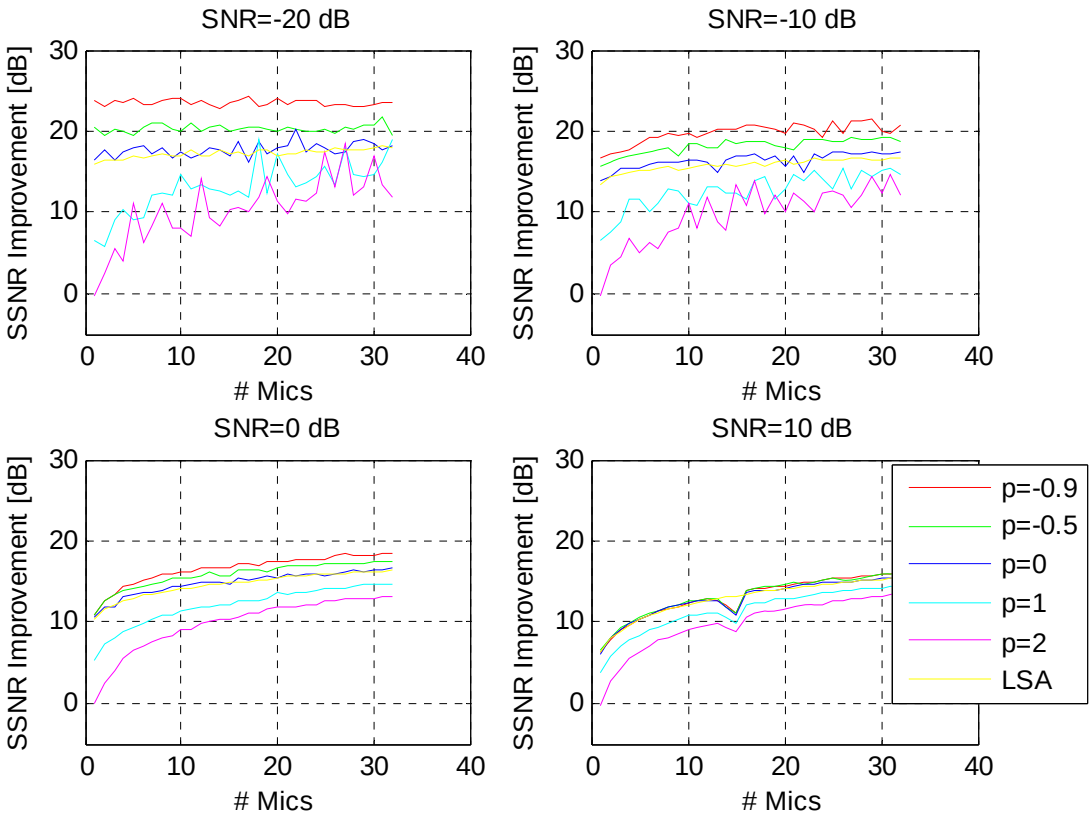

Figure 4-13 SSNR Improvements for Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Spectral Phase Estimation (Unity Attenuation Factors)

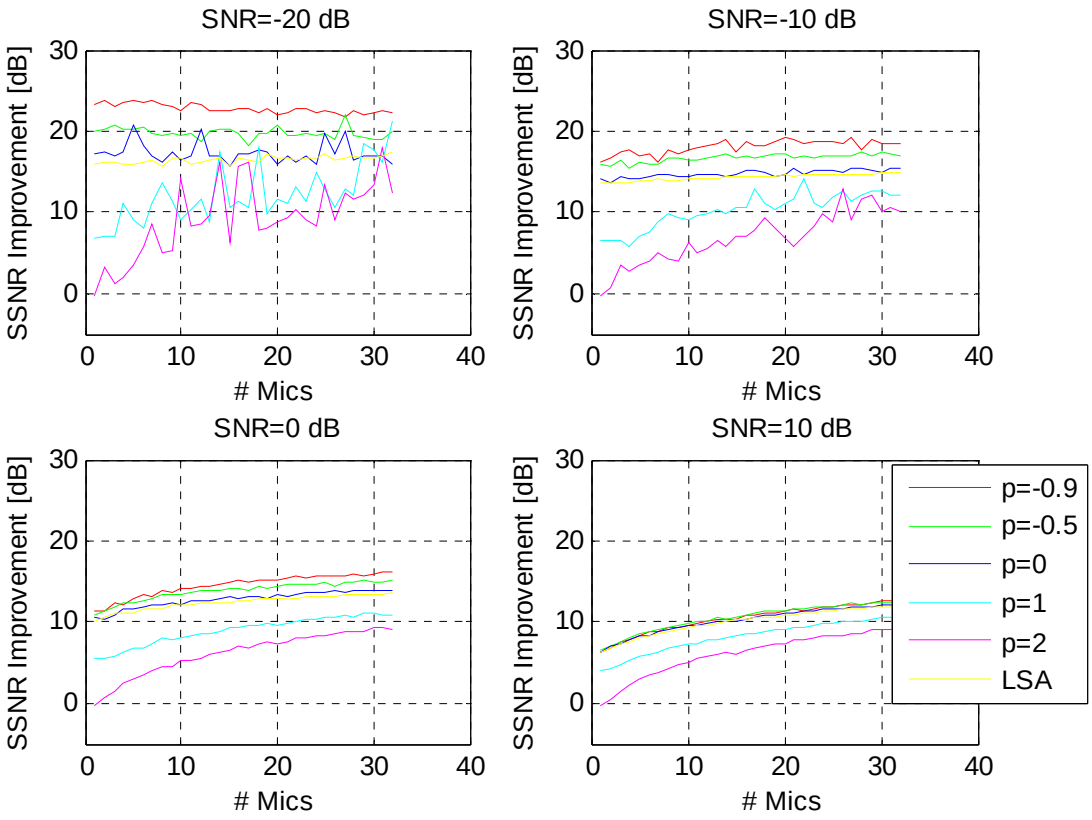

Figure 4-14 SSNR Improvements for Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Spectral Phase Estimation (Linear Attenuation Factors)

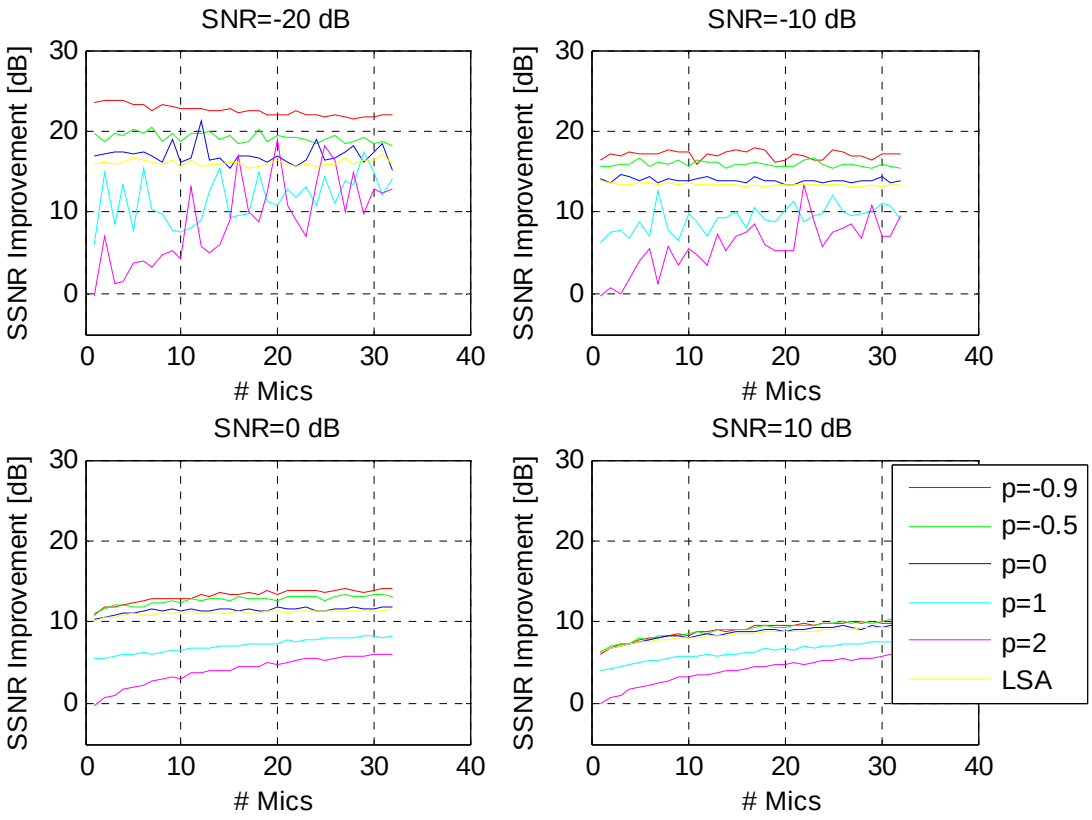

Figure 4-15 SSNR Improvements for Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Spectral Phase Estimation (Logarithmic Attenuation Factors)

In comparing the results with different attenuation factors, the WE and WCOSH estimators with spectral phase estimator both outperformed the STSA and LSA estimators with spectral phase estimator and time domain estimator for virtually all attenuation factor and input SNR cases by about 1-10 dB. As with the other estimators, the unity attenuation factors provided the best results for both the WE and WCOSH estimators with spectral phase estimator. In general, the WE estimator with spectral phase estimator had higher SSNR improvement over the WCOSH estimator with spectral phase estimator by 1-2 dB for most of the input SNR conditions for the optimal p values.

Figure 4-16 and Figure 4-17 show the time-series and spectrograms for the true source signal, noisy signal, and true source signal estimate.

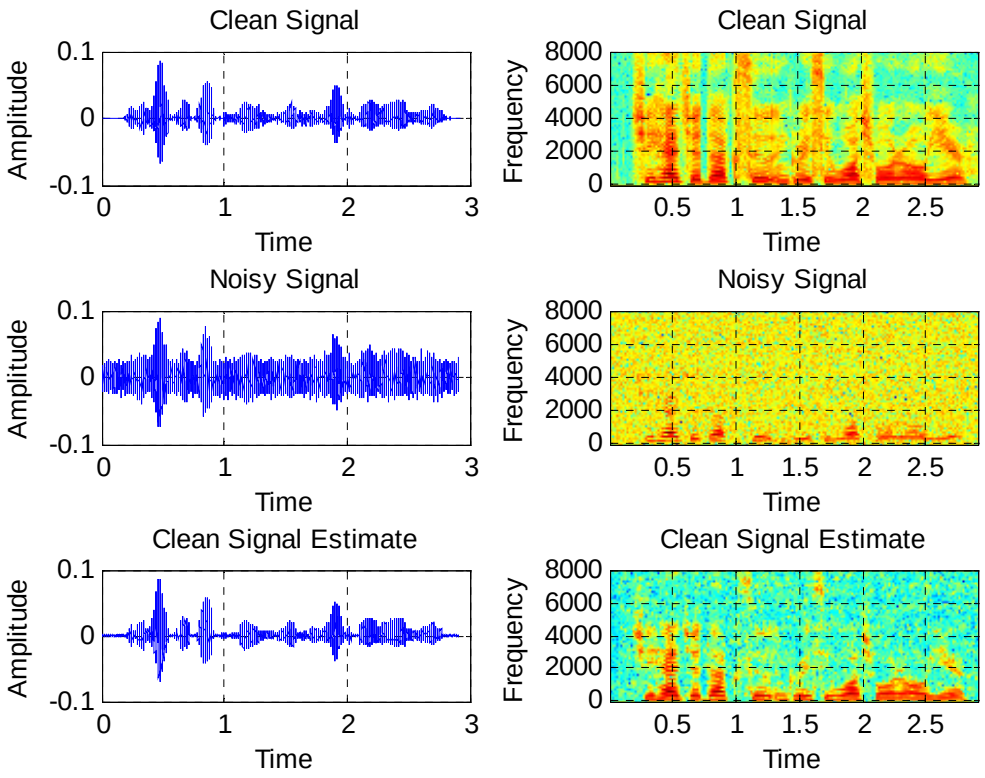

Figure 4-16 Time-Series and Spectrograms of Clean Signal, Noisy Signal, and Clean Signal Estimate for Weighted Euclidean (WE) Spectral Amplitude Estimation with Spectral Phase Estimation (Unity Attenuation Factors, 0 dB Input SNR, 32 Microphones)

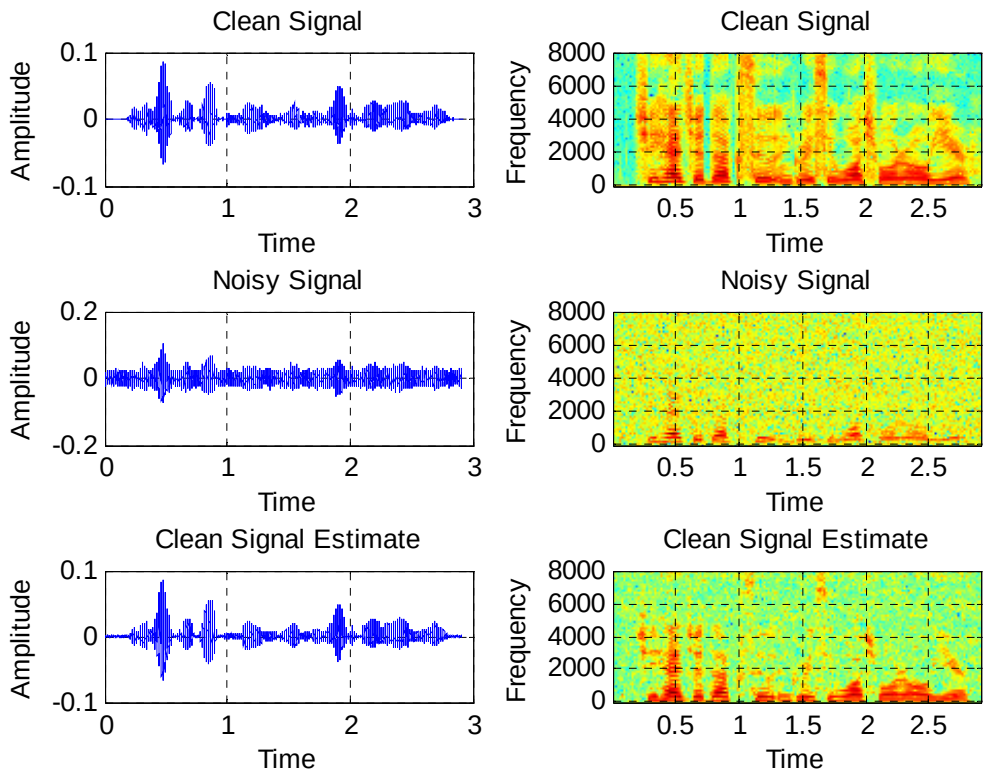

Figure 4-17 Time-Series and Spectrograms of Clean Signal, Noisy Signal, and Clean Signal Estimate for Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Spectral Phase Estimation (Unity Attenuation Factors, 0 dB Input SNR, 32 Microphones)

In contrast to the STSA and LSA estimators with spectral phase estimators in Figure 4-8 and Figure 4-9, the speech energies are much more well-defined in the spectrograms for the WE and WCOSH estimators with spectral phase estimator. Essentially, WE and WCOSH estimators are Bayesian estimators that emphasize spectral valleys (where noise is audible) more than the spectral peaks (where noise is inaudible and masked by the formants) and performed the best in terms of having less background noise and better speech quality. Overall, the perceptually-motivated spectral amplitude WE and WCOSH

estimators with spectral phase estimator provided even better enhancement results than the STSA and LSA estimators with spectral phase estimator.

4.3.1.4. Complex Real and Imaginary Component

The simulation results of 10 trial runs of the same sentence are shown for unity (Figure 4-18), linear (Figure 4-19), and logarithmic (Figure 4-20) attenuation factors as a function of increasing number of microphones.

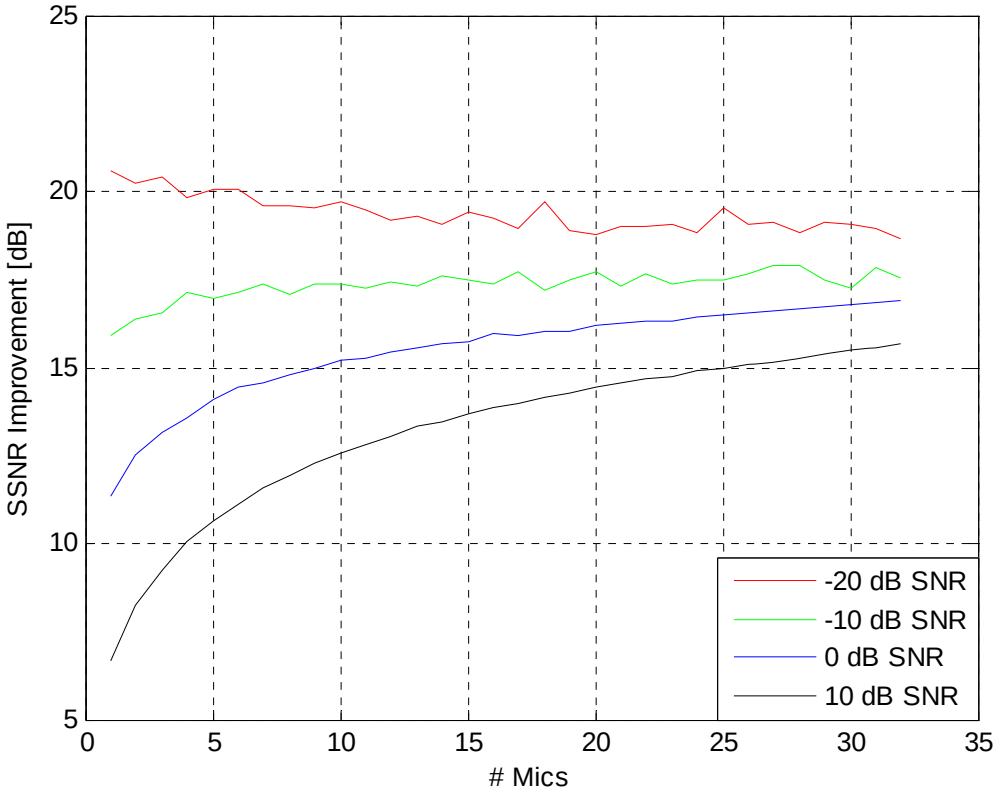

Figure 4-18 SSNR Improvements for Complex Real and Imaginary Spectral Component Estimation (Unity Attenuation Factors)

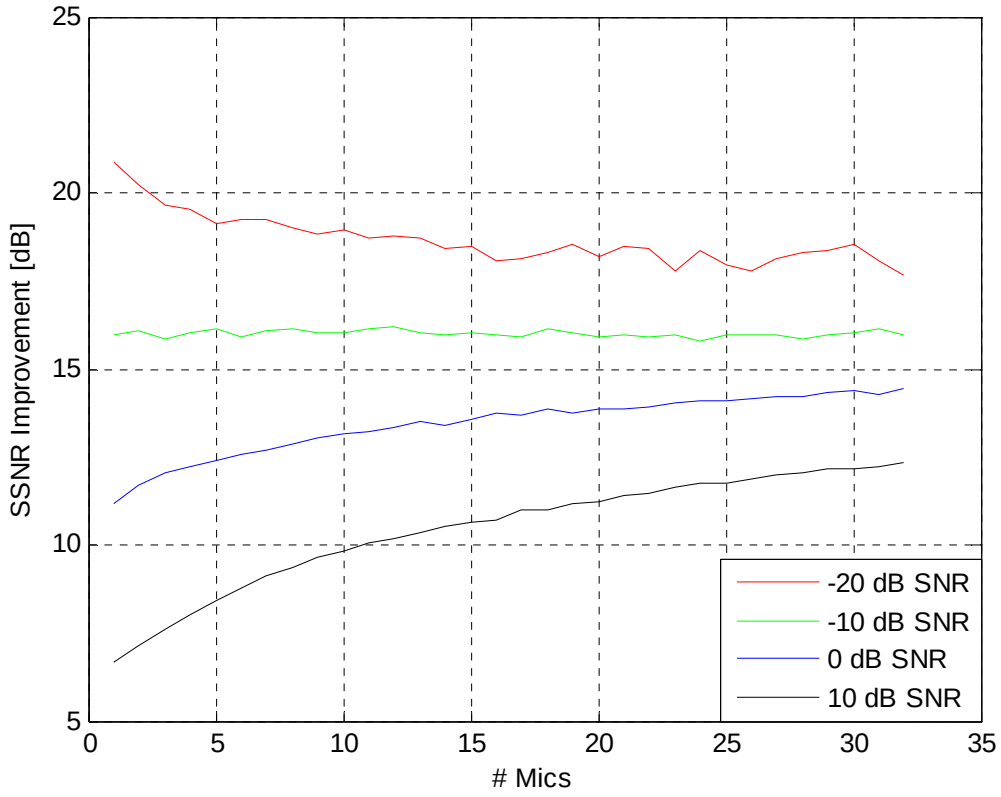

Figure 4-19 SSNR Improvements for Complex Real and Imaginary Spectral Component Estimation (Linear Attenuation Factors)

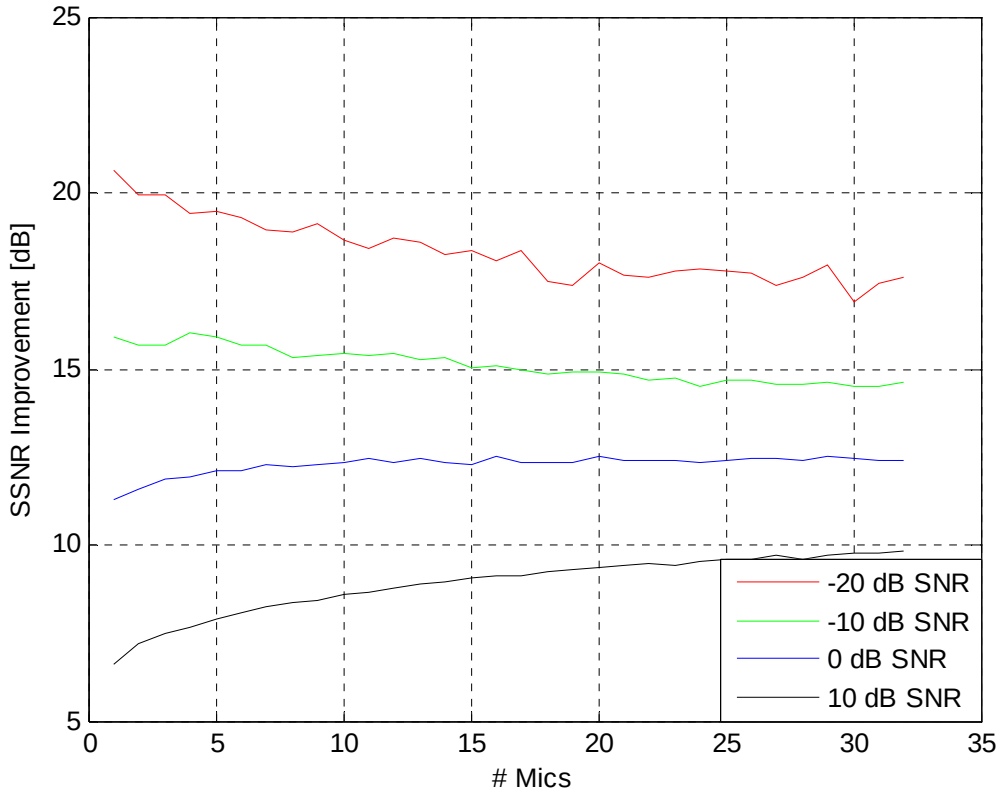

Figure 4-20 SSNR Improvements for Complex Real and Imaginary Spectral Component Estimation (Logarithmic Attenuation Factors)

The complex real and imaginary spectral component estimator outperformed the time domain estimator and STSA and LSA estimators with spectral phase estimator for all attenuation factors and input SNR cases by about 1-7 dB, except at -20 dB input SNR with the results actually being worse with additional microphones. Typically, the SSNR improvement was about 1 dB with -10 dB, 0 dB, and 10 dB input SNR and 5 dB and 1 dB at -20 dB for 1 microphone and 32 microphones. In comparison, the WCOSH and WE estimators with spectral phase estimator had better SSNR improvement results than the complex real and imaginary spectral component estimator by about 1-7 dB. Figure

4-21 show the time-series and spectrograms for the true source signal, noisy signal, and true source signal estimate.

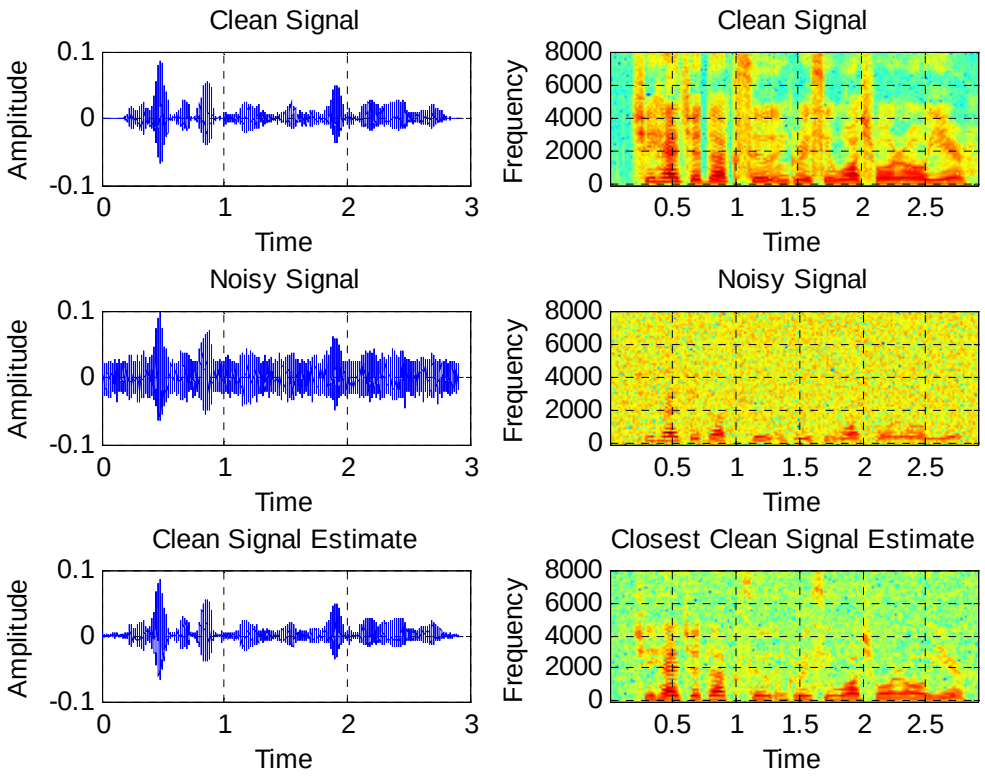

Figure 4-21 Time-Series and Spectrograms of Clean Signal, Noisy Signal, and Clean Signal Estimate for Complex Real and Imaginary Spectral Component Estimation (Unity Attenuation Factors, 0 dB Input SNR, 32 Microphones)

The complex real and imaginary spectral component estimator was able to capture speech energies and reduce background noise similar to the STSA and LSA estimators with spectral phase estimator but not quite as well as with the WCOSH and WE estimators with spectral phase estimator. Since the formulation of the frequency domain estimators can be done equivalently with the spectral amplitude and spectral phase and real and imaginary spectral components, the results should be fairly similar to each other. In

general, the complex real and imaginary spectral component estimator produced comparable SSNR improvement results and spectrograms to the STSA and LSA estimators with spectral phase estimator.

4.3.2. Spectral Phase Estimation

To demonstrate the benefit of the derived multichannel spectral phase estimator, simulations were performed and averaged over 10 trial runs of the same sentence using the multichannel STSA and multichannel LSA estimators with multichannel spectral phase estimator with unity attenuation factors. Results are illustrated in Figure 4-22, which show the SSNR improvement difference between the optimal multichannel spectral phase estimator and optimal single channel (noisy) spectral phase estimator.

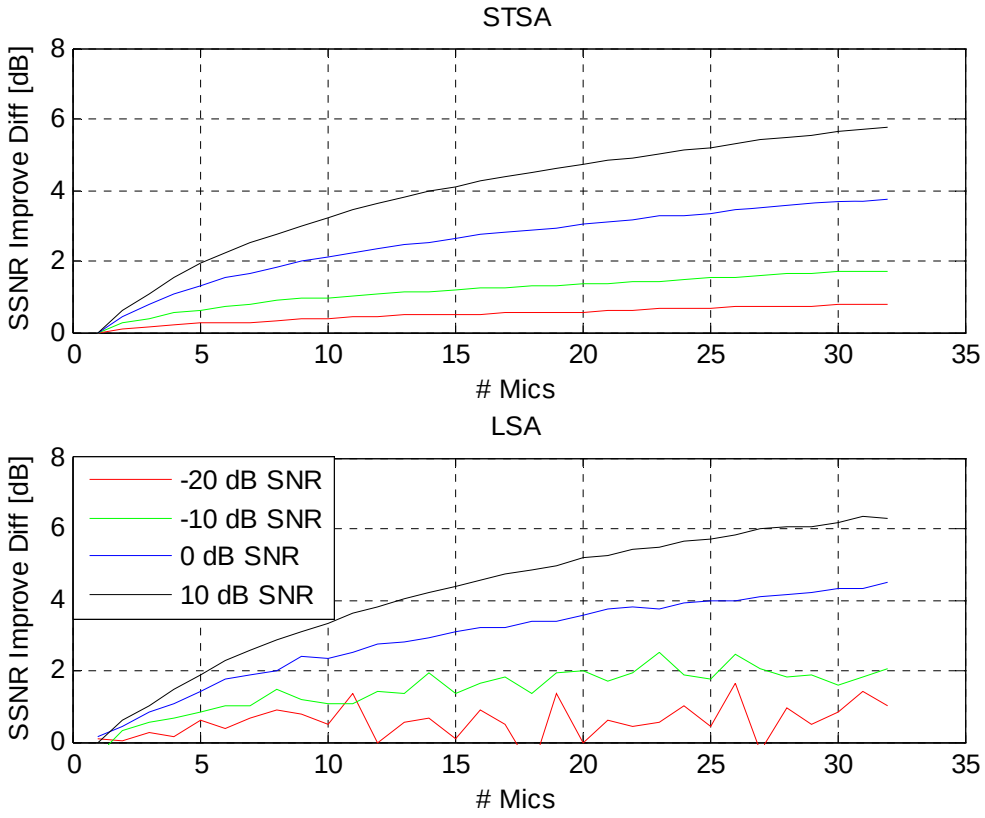

Figure 4-22 SSNR Improvement Difference between Multichannel Short-Time Spectral Amplitude (STSA) and Multichannel Log-Spectral Amplitude (LSA) Estimation with Multichannel Spectral Phase Estimation and Single Channel Short-Time Spectral Amplitude (STSA) and Single Channel Log-Spectral Amplitude (LSA) Estimation with Single Channel (Noisy) Spectral Phase Estimation (Unity Attenuation Factors)

The derived multichannel spectral phase estimator surpasses the single channel (noisy) single channel spectral phase estimator by a range of range of 0 dB to 5.8 dB (STSA estimator) and 0 dB to 6.2 dB (LSA estimator) across all of the input SNR levels, which increases with additional microphones. The greatest SSNR improvement differences appear for 32 microphones at the higher SNR levels of 0 dB (4.5 dB) and 10 dB (6.2 dB) with the LSA estimator. In general, it can be seen that the improvement due to the multichannel spectral phase estimator over the single channel (noisy) spectral phase

estimator constitutes a significant portion of the overall improvement obtained when using all of the available acoustic and spatial information from the noisy observations Y_i in the surrounding environment.

4.3.3. Time Alignment

Figure 4-23 illustrates the effects of artificial misalignment and corresponding automatic time alignment of a 32 channel configuration for the LSA estimator with spectral phase estimator using unity attenuation factors and 0 dB input SNR noisy signals.

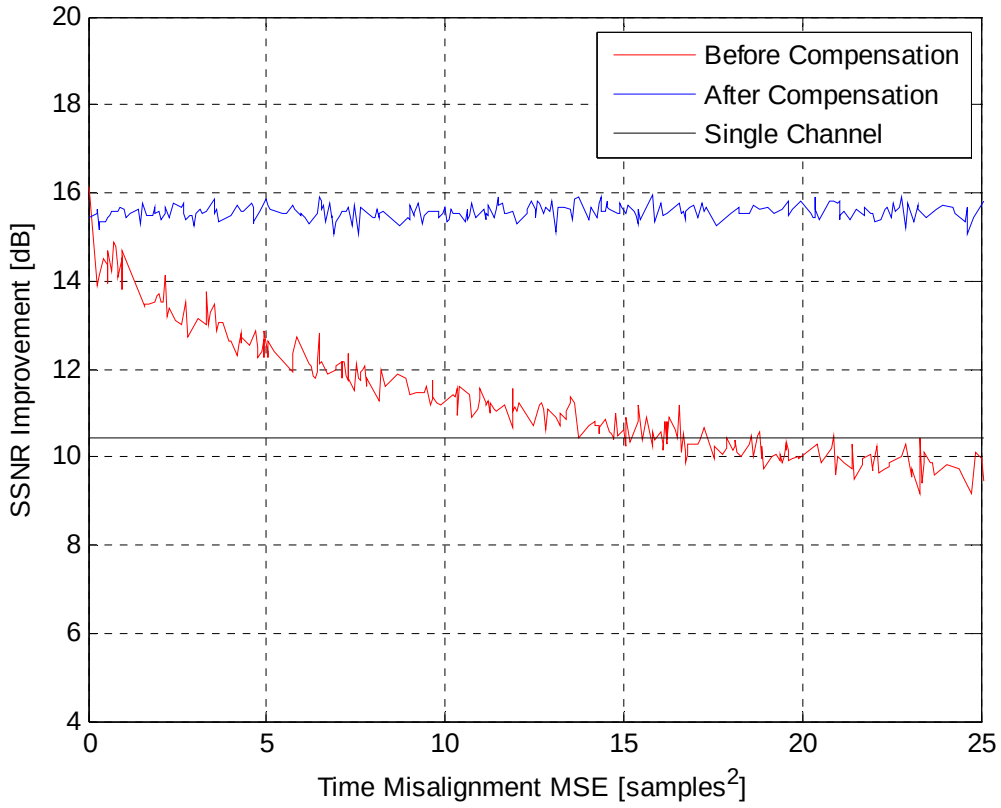

Figure 4-23 SSNR Improvement for 32 Microphones, 0 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

The degree of SSNR improvement is shown as a function of the amount of artificial misalignment measured by MSE. Enhancement decreased rapidly with an artificial misalignment of only 1 sample^2. With an artificial misalignment of only 18 samples^2, the SSNR improvements fell to already below the single channel LSA estimator with single channel spectral estimator of 10.42 dB. Consequently, the benefit of addition microphones vanishes for even a relatively small amount of artificial misalignment. It is clear that the success of the estimation relies directly on the ability to accurately time align the noisy signals y_i. The cross-correlation compensation method performed well

for time alignment with overall enhancement remaining steady at approximately 15.5 dB for artificial misalignment upwards of 25 samples^2, which is only slightly less than a 1 dB degradation in the results from perfect time alignment. Even after the cross-correlation compensation, the average remaining artificial misalignment was relatively minor for three of the four input SNR conditions: 2.9×10^7 samples^2 (-20 dB), 16.2 samples^2 (-10 dB), 14.4 samples^2 (0 dB), and 14.0 samples^2 (10 dB). With no overall trend suggesting a performance decrease as a function of initial artificial misalignment, it is possible to accurately ascertain time alignment.

4.3.4. Attenuation Factor Estimation

Figure 4-24 depicts the effects on SSNR improvements of the artificial error added to the true unity attenuation factors $c_i = 1$ of a 32 channel configuration for the LSA estimator with spectral phase estimator and 0 dB input SNR noisy signals.

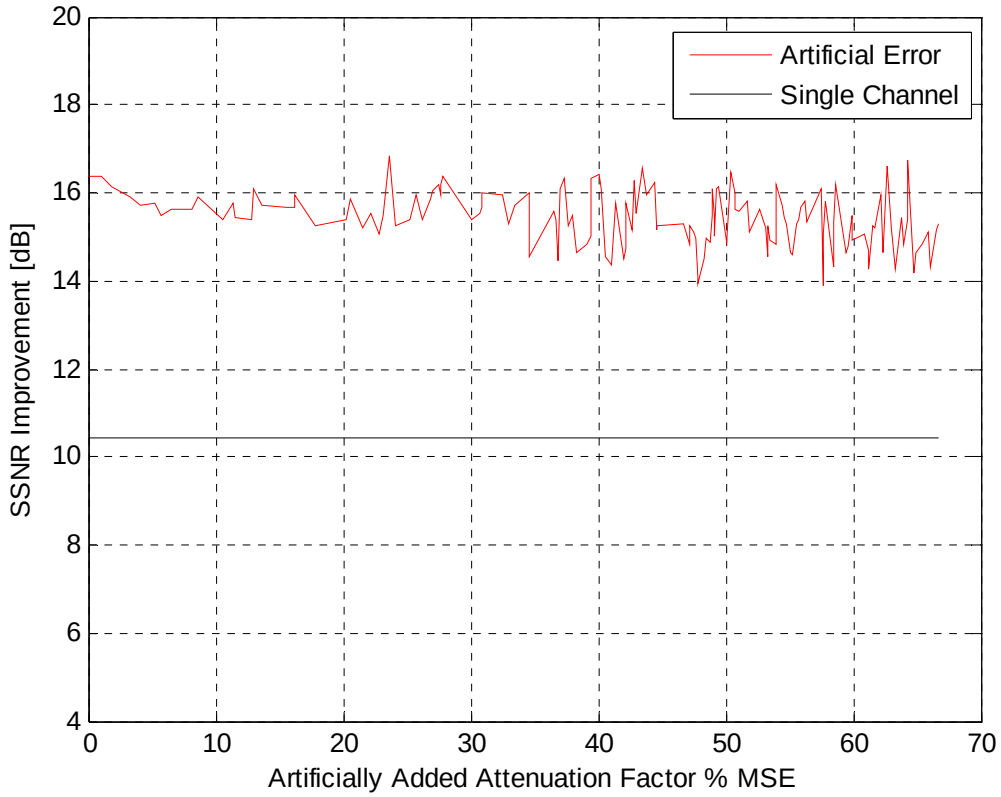

Figure 4-24 SSNR Improvement for 32 Microphones, 0 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

The results indicate that misestimation of the unity attenuation factors due to artificially added error caused a decrease in overall performance of about 1 dB at 5%-30% MSE. The small amounts of misestimation caused by the artificially added error did not have substantial impact on enhancement performance since even a wide range of 30%-70% MSE caused the overall enhancement to reach a steady 15.0 dB SSNR improvement, which is approximately a 1.5 dB decrease in the results with no artificially added error in the attenuation factors. Theoretically, the worst-case impact for attenuation factor misestimation would occur when a single channel has a dominatingly large attenuation

factor, which reduces performance to the single channel estimator applied to that particular channel. Without any artificially added error to the unity attenuation factors, the actual computed error in the estimated unity attenuation factors was approximately 2.58% (0 dB input SNR), 30.79% (-20 dB input SNR), 21.59% (-10 dB input SNR), and 0.24% (10 dB input SNR), which caused only a 1-2 dB degradation in the overall enhancement results. Therefore, there is very little performance decrease from misestimation of the unity attenuation factors due to any actual error or artificially added error.

4.4. Summary

The time domain and frequency domain statistical estimators provided significant gains in SSNR improvements with an increase in the number of microphones for distributed microphone speech enhancement. In the first set of results, enhancement, the frequency domain estimators had better performance than the simple time domain estimator. Between the unity, linear, and logarithmic attenuation factors, the unity attenuation factors always produced the best overall enhancement from the input SNR and varying number of microphones since each of the observations contain equally useful information. In contrast, the linear and then logarithmic attenuation factors had slightly less SSNR improvement with additional microphones because the single channel results had already reached nearly optimal performance. Table 4-7 displays a summary of the enhancement results for the various statistical estimators with unity, linear, and logarithmic attenuation factors.

SSNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	14.89	0.00	10.57	0.00	7.08
STSA + Estimated Phase	8.86	14.94	8.86	11.67	8.86	9.20
STSA + Noisy Phase	8.86	11.20	8.86	9.87	8.86	8.47
LSA + Estimated Phase	10.42	16.35	10.21	13.59	10.13	11.49
LSA + Noisy Phase	10.28	11.83	10.25	11.21	10.28	10.45
WE + Estimated Phase (p=-1.9)	**11.27**	**18.73**	**11.31**	**16.36**	**11.14**	**14.38**
WCOSH + Estimated Phase (p=-0.9)	10.97	18.44	11.29	16.15	10.91	14.09
Complex	11.35	16.90	11.15	14.45	11.30	12.37

Table 4-7 SSNR Improvement (Input SNR/SSNR = 0.0 dB/-7.6 dB)

Based on the enhancement results, the perceptually-motivated WE spectral amplitude estimator with spectral phase estimator outperformed all of the other estimators for the

optimal parameter of $p = -1.9$ because the estimator emphasized spectral valleys more than spectral peaks and had less remaining background noise and better overall speech quality.

In the second set of results, spectral phase estimation, the derived multichannel spectral phase estimator showed significant SSNR improvements over the single channel spectral phase estimator. Without the inclusion of the spectral phase estimator, the performance decreased linearly from 1 dB at 3% MSE to 3 dB at 7% MSE with SSNR improvement below the single channel baseline of 10 dB at 22% MSE. By including the multichannel spectral phase estimator, the performance exceeded the single channel (noisy) spectral phase estimator by as much as 6.5 dB for the best case of unity attenuation factors. Due to the STSA and LSA estimators and WE and WCOSH estimators requiring spectral phase information to reconstruct the enhanced signal, the multichannel spectral phase estimator was a vital aspect in improving the SSNR results with additional microphones since it utilized information about all of the noisy microphone observations, not simply information from the best microphone.

In the third set of results, time alignment, the performance of the statistical estimators seriously deteriorated with unsynchronized frames. For an artificial misalignment of simply 20 samples^2 with 32 microphones at 0 dB input SNR and unity attenuation factors, the enhancement results decreased from 16 dB to 9 dB, which is already 1.3 dB lower than the single channel enhancement results. By using cross-correlation to time align the noisy observations, there was only a reduction of 1 dB in SSNR improvement.

In the fourth set of experiments, attenuation factor estimation, the attenuation factors were artificially given random error on the known unity attenuation factors for 32 microphones and 0 dB input SNR. The amount of influence that artificial attenuation factor error had in the overall results for 32 microphone channels was negligible for the 0 dB and 10 dB cases, 1 dB degradation in the -10 dB case, and 3 dB degradation for the -20 dB case.

Overall, the success of the distributed speech enhancement estimators depends primarily on the assumption of independence of noise spectral components at each of the respective microphones, accurate time alignment of the noisy observations, and reasonable estimation of the attenuation factors.

CHAPTER 5 CONCLUSION

In this dissertation, the goal has been to generalize the speech enhancement work from single channel microphones, dual channel microphones, and microphone arrays for the purposes advancing the current state-of-the-art methods for distributed microphone speech enhancement. In order to realize these novel frameworks, the focus has been on developing and implementing robust and optimal time domain and frequency domain estimators for estimating the true source signal and measuring the SSNR performance improvement. By utilizing statistical estimation techniques and Gaussian distributions for the speech prior and noise likelihood models, the theoretical methods were derived for five basic classes of estimators: time domain, short-time spectral amplitude (STSA) with spectral phase, log-spectral amplitude (LSA) with spectral phase, perceptually-motivated spectral amplitude with spectral phase, and complex real and imaginary spectral component.

From the experimental work with different true source signal attenuation factors, the estimators all demonstrated significant gains in SSNR with an increase in the number of microphones, especially with the inclusion of the optimal multichannel spectral phase estimator. The recommendation here would be to use the perceptually-motivated spectral amplitude estimators with spectral phase estimator, namely the Weighted Euclidean (WE) and Weighted Cosh (WCOSH) estimators for approximately 10-15 microphones to offset overhead from the extra microphone costs, computation costs, and CPU time from the additional microphones. The statistical estimators have already shown tremendous promise with distributed microphone speech enhancement of noisy acoustic signals and could potentially be implemented in various consumer, industrial, and military

applications under severely noisy environments. Ultimately, the impact of this work is that it provides researchers, the general public, and law enforcement and national security agencies with methods for enhancing noisy speech collected by distributed microphones involving an assortment of target applications such as courtroom and meeting room transcriptions, broadcast news transcriptions, and speaker spotting, identification, and tracking systems.

5.1. Summary of Work

Optimal estimators have been developed and implemented in both the time domain and frequency domain for estimating the true source signal s, spectral amplitude A with spectral phase α, and complex real and imaginary component $S_{R,I}$ in a distributed microphone scenario. The distributed microphone speech enhancement system provided significant gains in SSNR improvements with an increase in the number of microphones and was demonstrated to be robust with respect to time alignment and estimation of the attenuation factors with corrupting uncorrelated and additive white Gaussian noise, particularly with the inclusion of the multichannel spectral phase estimator. Overall, the best enhancement results were from the perceptually-motivated spectral amplitude estimators with spectral phase frequency domain estimator: weighted cosh (WCOSH) and weighted Euclidean (WE).

5.2. Research Contributions

By utilizing statistical estimation techniques, the research has been able to extend the single channel microphone, dual channel microphone, and microphone array speech

enhancement methods into the distributed microphone scenario with five major classes of contributions:

1. Time Domain Estimation
2. Spectral Amplitude Estimation
3. Perceptually-Motivated Spectral Amplitude Estimation
4. Spectral Phase Estimation
5. Complex Real and Imaginary Spectral Component Estimation

Based on the experimental results, the statistical estimators are able to estimate the true source signal from noisy observations and improve the quality and intelligibility over the best case single channel results, particularly with the inclusion of the optimal multichannel spectral phase estimator.

5.3. Future Work

For future work, there are several directions for improving the distributed microphone speech enhancement estimators. While the dissertation explored uncorrelated and additive Gaussian white noise with estimation of the noise statistics involving simply an average of the initial frames of silence, the natural extension would be to utilize uncorrelated and additive real world non-Gaussian white noise (e.g., babble) from the NOISEX corpus [36] and employ time domain and frequency domain noise tracking estimators. It would also be possible to extend other single channel microphone speech enhancement estimators such as beta-order [37], alternative perceptually-motivated spectral amplitude [12], and additional complex real and imaginary complex [13] estimators. Instead of utilizing time-invariant attenuation factors for stationary sources,

the distributed microphone model could be modified to include time-variant attenuation factors to deal with moving sources in a noisy environment.

Besides investigating other speech enhancement statistical estimators, research needs to be invested towards developing and implementing novel algorithms with distributed microphones for robust and optimal speech recognition. There are numerous state-of-the-art methods for single channel feature enhancement [38-40], feature compensation [41], and model adaptation [42], but there are not yet any corresponding methods for distributed microphones. Table 5-1 shows the state-of-the-art methods for speech enhancement with the newly derived estimators from this work along with the state-of-the-art methods for speech recognition.

<table>
<tr><th>Method</th><th>Single Channel Microphones</th><th>Dual Channel Microphones</th><th>Microphone Arrays</th><th>Distributed Microphones</th></tr>
<tr><td rowspan="6">Speech Enhancement</td><td>Spectral Subtraction [9]</td><td rowspan="6">Adaptive Noise Cancellation [2]</td><td rowspan="3">Fixed Beamforming [3]</td><td>Time Domain</td></tr>
<tr><td>Wiener Filter [11]</td><td>Wiener Filter [10]</td></tr>
<tr><td>Short-Time Spectral Amplitude Estimation [5]</td><td>Short-Time Spectral Amplitude Estimation [7]</td></tr>
<tr><td>Log-Spectral Amplitude Estimation [6]</td><td rowspan="3">Adaptive Beamforming [3]</td><td>Log-Spectral Amplitude Estimation</td></tr>
<tr><td>Perceptually-Motivated Spectral Amplitude Estimation [12]</td><td>Perceptually-Motivated Spectral Amplitude Estimation</td></tr>
<tr><td>Complex Real and Imaginary Spectral Component Estimation [13]</td><td>Complex Real and Imaginary Spectral Component Estimation</td></tr>
<tr><td rowspan="2">Feature Enhancement</td><td>Filterbanks [38, 39]</td><td rowspan="2">None</td><td>LIMABEAM [43]</td><td rowspan="2">None</td></tr>
<tr><td>Cepstrals [40]</td><td>S-LIMABEAM [44]</td></tr>
<tr><td>Feature Compensation</td><td>CSM [41]</td><td>None</td><td>None</td><td>None</td></tr>
<tr><td>Model Adaptation</td><td>Phase-JAC/VTS [42]</td><td>None</td><td>None</td><td>None</td></tr>
</table>

Table 5-1 Standard Methods for Single Channel, Dual Channel Microphones, Microphone Arrays, and Distributed Microphone Speech Enhancement and Feature Enhancement, Feature Compensation, and Model Adaptation for Speech Recognition

In the next stage, the research work with distributed microphone will involve formulation and derivation of statistical estimation solutions in the feature domain (e.g., filterbank domain, cepstral domain) with innovative noise tracking algorithms for increasing the recognition accuracy of the estimated true source signal. Ultimately, the corresponding estimators will then be implemented and compared to the state-of-the-art single channel speech recognition estimators to determine the benefit of using additional microphones on recognition accuracy.

REFERENCES

[1] J. Polastre, R. Szewczyk, and A. Mainwaring, "Chapter 18: Analysis of Wireless Sensor Networks for Habitat Monitoring," in *Wireless Sensor Networks*. Norwell, MA: Kluwer Academic Publishers, 2004.

[2] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., and R. C. Goodlin, "Adaptive Noise Cancelling: Principles and Applications," *Proceedings of the IEEE*, vol. 63, pp. 1692-1716, 1975.

[3] M. Brandstein and D. Ward, *Microphone Arrays*. New York, NY: Springer-Verlag, 2001.

[4] I. A. McCowan, "Robust Speech Recognition using Microphone Arrays," Queensland University of Technology, 2001.

[5] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-32, pp. 1109-1121, 1984.

[6] Y. Ephraim and D. Malah, "Speech Enhancement using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 443-445, 1985.

[7] T. Lotter, C. Benien, and P. Vary, "Multichannel Direction-Independent Speech Enhancement Using Spectral Amplitude Estimation," *EURASIP Journal on Applied Signal Processing*, pp. 1147-1156, 2003.

[8] B. D. V. Veen and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," in *IEEE ASSAP Magazine*, 1988.

[9] S. F. Boll and D. C. Pulsipher, "Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancellation," 1980.

[10] S. Doclo and M. Moonen, "GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 50, pp. 2230-2244, 2002.

[11] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series, with Engineering Applications*. New York: Wiley, 1949.

[12] P. C. Loizou, "Speech Enhancement Based on Perceptually Motivated Bayesian Estimators of the Magnitude Spectrum," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 13, pp. 857-869, 2005.

[13] R. Martin, "Speech Enhancement Based on Minimum Mean-Square Error Estimation and Supergaussian Priors," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 13, pp. 845-856, 2005.

[14] H. L. v. Trees, *Detection, Estimation, and Modulation Theory*, vol. I. New York, NY: Wiley, 1968.

[15] H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed, 1994.

[16] "IEEE Recommended Practice for Speech Quality Measurements," *IEEE Transactions on Audio and Electroacoustics*, pp. 227-246, 1969.

[17] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 9, pp. 504-512, 2001.

[18] I. Cohen, "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement," *IEEE Signal Processing Letters*, vol. 9, pp. 12-15, 2002.
[19] I. Cohen, "Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 11, pp. 466-475, 2003.
[20] P. C. Loizou, *Speech Enhancement Theory and Practice*: CRC Press, 2007.
[21] S. F. Boll, "Suppression of Acoustic Noise Using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, pp. 113-120, 1979.
[22] M. Berouti, M. Schwartz, and J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," presented at Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1979.
[23] S. S. Haykin, *Adaptive Filter Theory, 4th Edition*. Upper Saddle River, NJ: Prentice-Hall, 2002.
[24] N. Wiener and E. Hopf, "On a Class of Singular Integral Equations," presented at Proceedings Russian Academic Mathematics Physics, 1931.
[25] I. S. Gradshteyn and Z. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th Edition ed. New York: Academic, 1994.
[26] I. S. Gradshteyn and Z. M. Ryzhik, *Table of Integrals, Series, and Products*. New York City, New York: Academic, 1980.
[27] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-33, pp. 443-445, 1985.
[28] J. J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," in *IEEE Signal Processing Magazine*, vol. 9, 1992, pp. 14-37.
[29] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-24, pp. 320-327, 1976.
[30] O. L. Frost, "An algorithm for linear constrained adaptive beamforming," *Proceedings of the IEEE*, vol. 60, pp. 926-935, 1972.
[31] L. J. Griffiths and C. W. Jim, "An Alternative Approach to Linear Constrained Adaptive Beamforming," *IEEE Transactions on Antennas Propagation*, vol. AP-30, pp. 27-34, 1982.
[32] D. Middleton, *Introduction to Statistical Communication Theory*. New York: McGraw-Hill, 1960.
[33] M. B. Trawicki and M. T. Johnson, "Optimal Distributed Microphone Phase Estimation," presented at International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Taipei, Taiwan, R.O.C., 2009.
[34] J. Garofolo, L. Lamel, and W. Fisher, "TIMIT Acoustic-Phonetic Continuous Speech Corpus." Linguistic Data Consortium, 1993.
[35] L. E. Kinsler, *Fundamentals of Acoustics*, 4th ed: John Wiley & Sons, Inc., 1999.
[36] A. Varga and H. J. M. Steeneken, "Assessment for Automatic Speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems," *Speech Communication*, vol. 12, pp. 247-251, 1993.

[37] C. H. You, S. N. Koh, and S. Rahardja, "Beta-Order MMSE Spectral Amplitude Estimation for Speech Enhancement," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 13, 2005.

[38] L. Deng, J. Droppo, and A. Acero, "Estimating Cepstrum of Speech Under the Presence of Noise Using a Joint Prior of Static and Dynamic Features," *IEEE Transactions on Speech and Audio Processing*, vol. 12, pp. 218-233, 2004.

[39] L. Deng, J. Droppo, and A. Acero, "Enhancement of Log Mel Power Spectra of Speech using a Phase-Sensitive Model of the Acoustic Environment and Sequential Estimation of the Corrupting Noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 12, 2004.

[40] D. Yu, L. Deng, J. Droppo, J. Wu, Y. Gong, and A. Acero, "Robust Speech Recognition Using a Cepstral Minimum-Mean-Square-Error-Motivated Noise Suppressor," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, pp. 1061-1070, 2008.

[41] H. K. Kim and R. C. Rose, "Cepstrum-Domain Acoustic Feature Compensation Based on Decomposition of Speech and Noise for ASR in Noisy Environments," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 435-446, 2003.

[42] J. Li, L. Deng, D. Yu, Y. Gong, and A. Acero, "HMM Adaptation using a Phase-Sensitive Acoustic Distortion Model for Environment-Robust Speech Recognition," presented at International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2008.

[43] M. L. Seltzer, B. Raj, and R. M. Stern, "Likelihood-Maximizing Beamforming for Robust Hands-Free Speech Recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 12, pp. 489-498, 2004.

[44] M. L. Seltzer and R. M. Stern, "Subband Likelihood-Maximizing Beamforming for Speech Recognition in Reverberant Environments," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, pp. 2109-2121, 2006.

APPENDIX A TIME DOMAIN ESTIMATOR

In this appendix, the minimum mean-square error time domain estimator is derived for distributed multichannel signals. To facilitate construction of the estimator $\hat{s}$, the mean μ_{s_i} is computed for each microphone i from (3.11). Since the exponent term inside the exponential is quadratic in nature, (3.11) is still a valid Gaussian distribution. Through the expansion of the exponential term and collecting the square s^2, linear s, and C constant terms, the summation term in (3.11) is rewritten as

$$\begin{aligned}
&\sum_{i=1}^{M} \frac{\left(s-\mu_{s_i}\right)^2}{\sigma_{s_i}^2} \\
&= \frac{s^2-2\mu_{s_1}s+\mu_{s_1}}{\sigma_{s_1}^2} + \frac{s^2-2\mu_{s_2}s+\mu_{s_2}}{\sigma_{s_2}^2} + \ldots + \frac{s^2-2\mu_{s_M}s+\mu_{s_M}}{\sigma_{s_M}^2} \\
&= \left[\sigma_{s_2}^2\sigma_{s_3}^2\cdots\sigma_{s_M}^2 + \sigma_{s_1}^2\sigma_{s_3}^2\cdots\sigma_{s_M}^2 + \ldots + \sigma_{s_1}^2\sigma_{s_2}^2\cdots\sigma_{s_{M-1}}^2\right]s^2 \\
&-2\left[\mu_1\left(\sigma_{s_2}^2\sigma_{s_3}^2\cdots\sigma_{s_M}^2\right) + \mu_{s_2}\left(\sigma_{s_1}^2\sigma_{s_3}^2\cdots\sigma_{s_M}^2\right) + \ldots + \mu_{s_M}\left(\sigma_{s_1}^2\sigma_{s_2}^2\cdots\sigma_{s_{M-1}}^2\right)\right]s \\
&+\left[\mu_{s_1}^2\left(\sigma_{s_2}^2\sigma_{s_3}^2\cdots\sigma_{s_M}^2\right) + \mu_{s_2}^2\left(\sigma_{s_1}^2\sigma_{s_3}^2\cdots\sigma_{s_M}^2\right) + \ldots + \mu_{s_M}^2\left(\sigma_{s_1}^2\sigma_{s_2}^2\cdots\sigma_{s_{M-1}}^2\right)\right]
\end{aligned} \tag{A.1}$$

or

$$\sum_{i=1}^{M} \frac{\left(s-\mu_{s_i}\right)^2}{\sigma_{s_i}^2} = As^2 - 2Bs + C, \tag{A.2}$$

where

$$A = \left(\sum_{i=1}^{M}\prod_{\substack{j=1\\ j\neq i}}^{M}\sigma_{s_j}^2\right), \tag{A.3}$$

$$B = -2\left(\sum_{i=1}^{M}\mu_{s_i}\prod_{\substack{j=1\\ j\neq i}}^{M}\sigma_{s_j}^2\right), \tag{A.4}$$

and

$$C = \sum_{i=1}^{M} \mu_{s_i}^2 \prod_{\substack{j=1 \\ j \neq i}}^{M} \sigma_{s_j}^2 \,. \tag{A.5}$$

In order to write the summation in (A.1) as a single term $\frac{\left(s - \mu_s\right)^2}{\sigma_s^2}$, the mathematical technique of completing the square is utilized to produce the mean

$$\mu_s = -\frac{B}{2A} \tag{A.6}$$

and variance

$$\sigma_s^2 = \frac{1}{A}. \tag{A.7}$$

By substituting (A.3) and (A.4) in (A.6), the closed-form solution $\hat{s}$ is given in (3.19) as

$$\mu_s = \hat{s} = \sum_{i=1}^{M} w_i y_i \tag{A.8}$$

with corresponding weights (3.20) as

$$w_i = \frac{\sigma_s}{\sqrt{\sigma_{y_i}^2 - \sigma_{n_i}^2}} \frac{\prod_{\substack{j=1 \\ j \neq i}}^{M} \sigma_{s_j}^2}{\sum_{i=1}^{M} \prod_{\substack{j=1 \\ j \neq i}}^{M} \sigma_{s_j}^2}. \tag{A.9}$$

APPENDIX B SHORT-TIME SPECTRAL AMPLITUDE ESTIMATOR

In this appendix, the minimum mean-square error short-time spectral amplitude estimator is derived for distributed multichannel signals. After substitution of the statistical models in (3.21) and (3.23), the result from (3.24) is

$$\hat{A}_{STSA} = \frac{\int_0^{\infty} A^2 \exp\left(-\frac{A^2}{\sigma_S^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha dA}{\int_0^{\infty} A \exp\left(-\frac{A^2}{\sigma_S^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha dA}. \tag{B.1}$$

The integration over the spectral phase α is performed by expansion of the term $\left|Y_i - c_i A e^{j\alpha}\right|^2 = \left(Y_i - c_i A e^{j\alpha}\right)_R^2 + \left(Y_i - c_i A e^{j\alpha}\right)_I^2$ and extracting the constants from the integral as

$$\begin{aligned} &\int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha \\ &= \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i\right|^2 + c_i^2 A^2}{\sigma_{N_i}^2}\right) \int_0^{2\pi} \exp\left(a\cos\alpha + b\sin\alpha\right) d\alpha \end{aligned}, \tag{B.2}$$

where

$$a = \sum_{i=1}^{M} \frac{2c_i A}{\sigma_{N_i}^2} \mathrm{Re}\left(Y_i\right) \tag{B.3}$$

and

$$b = \sum_{i=1}^{M} \frac{2c_i A}{\sigma_{N_i}^2} \mathrm{Im}\left(Y_i\right). \tag{B.4}$$

From trigonometric identities, the sum of cosine and sine terms with different amplitudes and the same phase is written as

$$a\cos\alpha + b\sin\alpha = \sqrt{a^2 + b^2}\,\cos\left(\alpha - \arctan\left(\frac{b}{a}\right)\right), \tag{B.5}$$

where

$$\sqrt{a^2 + b^2} = 2A\left|\sum_{i=1}^{M}\frac{c_i Y_i}{\sigma_{N_i}^2}\right|. \tag{B.6}$$

Since the integral in (B.2) for the spectral phase α is over one full period, the spectral phase shift of $\arctan\left(\frac{b}{a}\right)$ is removed from (B.5). By means of equation 8.431.1 in [25], the integral in (B.2) is rewritten as

$$\int_0^{2\pi}\exp\left(a\cos\alpha + b\sin\alpha\right)d\alpha = 2\pi I_0\left(2A\left|\sum_{i=1}^{M}\frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right), \tag{B.7}$$

which reduces (B.1) to

$$\hat{A}_{STSA} = \frac{\int_0^{\infty} A^2\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\sum_{i=1}^{M}\frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right)dA}{\int_0^{\infty} A\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\sum_{i=1}^{M}\frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right)dA}. \tag{B.8}$$

Through substitution of equations 8.406.3 and 6.631.1 in [26] and [25], the closed-form solution $\hat{A}_{STSA}$ is given in (3.25) as

$$\hat{A}_{STSA} = \frac{\Gamma(1.5)}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \frac{{}_1F_1\left(\frac{3}{2};1;v\right)}{{}_1F_1(1;1;v)}. \tag{B.9}$$

APPENDIX C LOG-SPECTRAL AMPLITUDE ESTIMATOR

In this appendix, the minimum mean-square error log-spectral amplitude estimator is derived for distributed multichannel signals. After substitution of (3.21) and (3.23), (3.33) is expressed as

$$\Phi_{Z|Y_1,\ldots,Y_M}(\mu)=\frac{\int_0^{\infty}A^{\mu+1}\exp\left(-\frac{A^2}{\sigma_S^2}\right)\int_0^{2\pi}\exp\left(-\sum_{i=1}^{M}\frac{\left|Y_i-c_iAe^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right)d\alpha dA}{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_S^2}\right)\int_0^{2\pi}\exp\left(-\sum_{i=1}^{M}\frac{\left|Y_i-c_iAe^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right)d\alpha dA}. \quad \text{(C.1)}$$

The integration over the spectral phase α is performed exactly as in APPENDIX B. By employing (B.2)-(B.7) from APPENDIX B, (C.1) is written as

$$\Phi_{Z|Y_1,\ldots,Y_M}(\mu)=\frac{\int_0^{\infty}A^{\mu+1}\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\sum_{i=1}^{M}\frac{c_iY_i}{\sigma_{N_i}^2}\right|\right)dA}{\int_0^{\infty}A\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\sum_{i=1}^{M}\frac{c_iY_i}{\sigma_{N_i}^2}\right|\right)dA}. \quad \text{(C.2)}$$

Through application of equations 8.406.3 and 6.631.1 in [26] and [25], the closed-form solution $\Phi_{Z|Y_1,\ldots,Y_M}(\mu)$ is established in (3.34) as

$$\Phi_{Z|Y_1,\ldots,Y_M}(\mu)=\frac{\Gamma\left(\frac{\mu}{2}+1\right)}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}}{}_1F_1\left(-\frac{\mu}{2};1;-\nu\right). \quad \text{(C.3)}$$

The differentiation of (3.34) with respect to μ results in three derivative terms that are written as

$$
\begin{aligned}
E\left[Z \left| Y_1,...,Y_M \right. \right] &= \frac{d}{d\mu}\left[\Phi_{Z|Y_1,...,Y_M}(\mu) \right]\Big|_{\mu=0} \\
&= \left[\frac{d}{d\mu}\left(\Gamma\left(\frac{\mu}{2}+1 \right) \right) \frac{1}{\left(\frac{1}{\lambda} \right)^{\frac{\mu}{2}}} {}_1F_1\left(-\frac{\mu}{2};1;-v \right) \right]\Big|_{\mu=0} \\
&+ \left[\Gamma\left(\frac{\mu}{2}+1 \right) \frac{d}{d\mu}\left(\frac{1}{\left(\frac{1}{\lambda} \right)^{\frac{\mu}{2}}} \right) {}_1F_1\left(-\frac{\mu}{2};1;-v \right) \right]\Big|_{\mu=0} \\
&+ \left[\Gamma\left(\frac{\mu}{2}+1 \right) \frac{1}{\left(\frac{1}{\lambda} \right)^{\frac{\mu}{2}}} \frac{d}{d\mu}\left({}_1F_1\left(-\frac{\mu}{2};1;-v \right) \right) \right]\Big|_{\mu=0}
\end{aligned}
\tag{C.4}
$$

and evaluated at $\mu = 0$. The derivative of the first term is evaluated exactly as in [6] using

$$
\frac{d}{d\mu}\left(\Gamma\left(\frac{\mu}{2}+1 \right) \right) = \Gamma\left(\frac{\mu}{2}+1 \right) \frac{d}{d\mu}\left(\ln\left(\Gamma\left(\frac{\mu}{2}+1 \right) \right) \right). \tag{C.5}
$$

Through the series expansion given by equation 8.342.1 in [26], the last term in (C.5) is rewritten as

$$
\ln\left(\Gamma\left(\frac{\mu}{2}+1 \right) \right) = -c\frac{\mu}{2} + \sum_{i=2}^{\infty} \frac{(-\mu)^r}{2^r r} \alpha_r \,, \tag{C.6}
$$

where $|\mu| < 2$, c is Euler's constant, and

$$
\alpha_r \triangleq \sum_{n=1}^{\infty} \frac{1}{n^r} \,. \tag{C.7}
$$

By differentiating (C.6) term-by-term and evaluating (C.5) at $\mu = 0$, the derivative of the first term in (C.4) is given as

$$\frac{d}{d\mu}\left[\Gamma\left(\frac{\mu}{2}+1\right)\right]\Bigg|_{\mu=0} = -\frac{c}{2}. \tag{C.8}$$

The derivative of the second term $\left(\frac{1}{\lambda}\right)^{-\frac{\mu}{2}}$ in (C.4) is computed in a straightforward manner by rewriting it in exponential form and evaluating at $\mu = 0$ as

$$\frac{d}{d\mu}\left[\frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{\mu}{2}}}\right]\Bigg|_{\mu=0} = \frac{d}{d\mu}\left[e^{\frac{1}{2}\mu\ln(\lambda)}\right]\Bigg|_{\mu=0} = \frac{1}{2}\ln(\lambda). \tag{C.9}$$

For the computation of the third term, the confluent hypergeometric function ${}_1F_1\left(-\frac{\mu}{2};1;-v\right)$ is differentiated through its series expansion from equation 9.210.1 in [25] as

$${}_1F_1(a;b;x) = \sum_{r=0}^{\infty}\frac{(a)_r}{(c)_r}\frac{x^r}{r!}, \tag{C.10}$$

where $(a)_r = 1\cdot a\cdot(a+1)\cdot\ldots\cdot(a+r-1)$ with $(a)_0 \triangleq 1$. By differentiating (C.10) term-by-term and evaluating at $\mu = 0$, the derivative is given as

$$\frac{d}{d\mu}\left[{}_1F_1\left(-\frac{\mu}{2};1;-v\right)\right]\Bigg|_{\mu=0} = -\frac{1}{2}\sum_{r=1}^{\infty}\frac{(-v)^r}{r!}\frac{1}{r}. \tag{C.11}$$

By combining the three derivative results in (C.8), (C.9), and (C.11), (C.4) reduces to

$$
\begin{aligned}
E\left[Z\left|Y_1,...,Y_M\right.\right] &= \frac{d}{d\mu}\left[\Phi_{Z|Y_1,...,Y_M}(\mu)\right]\Big|_{\mu=0} \\
&= \left(-\frac{c}{2}\right){}_1F_1(0;1;-v) + \ln\left(\sqrt{\lambda}\right){}_1F_1(0;1;-v) + \left(-\frac{1}{2}\sum_{r=1}^{\infty}\frac{(-v)^r}{r!}\frac{1}{r}\right), \\
&= -\frac{1}{2}\left[c + \sum_{r=1}^{\infty}\frac{(-v)^r}{r!}\frac{1}{r}\right] + \frac{1}{2}\ln(\lambda)
\end{aligned}
\tag{C.12}
$$

where ${}_1F_1(0;1;-v) = 1$. From equations 8.211.1 and 8.214.1 in [26], (C.12) is rewritten as

$$
\begin{aligned}
E\left[Z\left|Y_1,...,Y_M\right.\right] &= -\frac{1}{2}\left[-\int_v^{\infty}\frac{e^{-t}}{t}dt - \ln(v)\right] + \frac{1}{2}\ln(\lambda) \\
&= -\frac{1}{2}\ln\left(\frac{1}{\lambda}\right) + \frac{1}{2}\left[\int_v^{\infty}\frac{e^{-t}}{t}dt + \ln(v)\right]
\end{aligned}.
\tag{C.13}
$$

Consequently, the final closed-form solution of $\hat{A}_{LSA}$ after exponentiation is written as in (3.36) as

$$
\hat{A}_{LSA} = \left(\frac{v}{\frac{1}{\lambda}}\right)^{\frac{1}{2}} \exp\left(\frac{1}{2}\int_v^{\infty}\frac{e^{-t}}{t}dt\right). \tag{C.14}
$$

APPENDIX D WE SPECTRAL AMPLITUDE ESTIMATOR

In this appendix, the perceptually-motivated weighted Euclidean (WE) spectral amplitude estimator is derived for distributed multichannel signals. By substitution of the statistical models in (3.21) and (3.23), (3.41) is written as

$$\hat{A}_{WE} = \frac{\int_0^{\infty} A^{p+2} \exp\left(-\frac{A^2}{\sigma_S^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha dA}{\int_0^{\infty} A^{p+1} \exp\left(-\frac{A^2}{\sigma_S^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha dA}. \tag{D.1}$$

As in APPENDIX B, the spectral phase α is integrated out from the both of the inner integrals as

$$\hat{A}_{WE} = \frac{\int_0^{\infty} A^{p+2} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\sum_{i=1}^{M} \frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right) dA}{\int_0^{\infty} A^{p+1} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\sum_{i=1}^{M} \frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right) dA}, \tag{D.2}$$

where $\frac{1}{\lambda}$ is defined in (3.26). By utilizing 8.406.3 and 6.631.1 in [26] and [25], (D.2) is given in terms of the gamma function $\Gamma(\bullet)$ and confluent hypergeometric function ${}_1F_1(\bullet;\bullet;\bullet)$ described by 9.210 in [25] as

$$\hat{A}_{WE} = \frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+1\right)} \frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \frac{{}_1F_1\left(\frac{p+3}{2};1;z\right)}{{}_1F_1\left(\frac{p+2}{2};1;z\right)}, \tag{D.3}$$

where

$$\frac{1}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} = \left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M} \xi_i} \right)^{\frac{1}{2}} \tag{D.4}$$

with $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$. From (D.3) and (D.4), the final closed-form solution $\hat{A}_{WE}$ is given in (3.43) as

$$\hat{A}_{WE} = \frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+1\right)} \left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M} \xi_i} \right)^{\frac{1}{2}} \frac{{}_1F_1\left(-\left(\frac{p+1}{2}\right);1;-z\right)}{{}_1F_1\left(-\frac{p}{2};1;-z\right)} \tag{D.5}$$

with free parameter $p_{WE} > 2$.

APPENDIX E WCOSH SPECTRAL AMPLITUDE ESTIMATOR

In this appendix, the perceptually-motivated weighted cosh (WCOSH) spectral amplitude estimator is derived for distributed multichannel signals. From substitution of the statistical models in (3.21) and (3.23), (3.44) is written as

$$\hat{A}^2_{WCOSH} = \frac{\int_0^{\infty} A^{p+2} \exp\left(-\frac{A^2}{\sigma_S^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha dA}{\int_0^{\infty} A^{p} \exp\left(-\frac{A^2}{\sigma_S^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha dA}. \quad \text{(E.1)}$$

After integrating out the spectral phase α from the both of the inner integrals as in APPENDIX B, (E.1) is given as

$$\hat{A}^2_{WCOSH} = \frac{\int_0^{\infty} A^{p+2} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\sum_{i=1}^{M} \frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right) dA}{\int_0^{\infty} A^{p} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A\left|\sum_{i=1}^{M} \frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right) dA}, \quad \text{(E.2)}$$

where $\frac{1}{\lambda}$ is defined in (3.26). Through 8.406.3 and 6.631.1 in [26] and [25], (E.2) is given in terms of the gamma function $\Gamma(\bullet)$ and confluent hypergeometric function ${}_1F_1(\bullet;\bullet;\bullet)$ described by 9.210 in [25] as

$$\hat{A}^2_{WCOSH} = \frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+\frac{1}{2}\right)} \frac{1}{\frac{1}{\lambda}} \frac{{}_1F_1\left(\frac{p+3}{2};1;z\right)}{{}_1F_1\left(\frac{p+1}{2};1;z\right)}, \quad \text{(E.3)}$$

where

$$\frac{1}{\frac{1}{\lambda}} = \frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i} \tag{E.4}$$

using $\sigma_{S_i}^2 = c_i^2\sigma_S^2$. As a result of (E.3) and (E.4), the closed-form solution of $\hat{A}_{WCOSH}$ is given in (3.44) as

$$\hat{A}_{WCOSH} = \sqrt{\frac{\Gamma\left(\frac{p}{2}+\frac{3}{2}\right)}{\Gamma\left(\frac{p}{2}+\frac{1}{2}\right)}\left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i}\right)\frac{{}_1F_1\left(-\left(\frac{p+1}{2}\right);1;-z\right)}{{}_1F_1\left(-\left(\frac{p-1}{2}\right);1;-z\right)}} \tag{E.5}$$

with free parameter $p_{WCOSH} > -1$.

APPENDIX F SPECTRAL PHASE ESTIMATOR

In this appendix, the minimum mean-square error spectral phase estimator is derived for distributed multichannel signals. After expanding the terms in the expectation with Euler's identity conditioned on the noisy spectral coefficients $\{Y_1,...,Y_M\}$, (3.46) is rewritten as

$$\begin{aligned}&\min_{g,\rho} E\left[\left|e^{j\alpha}-g\right|^2\left|Y_1,...,Y_M\right.\right]+\rho\left(|g|-1\right)\\&=\min_{g,\rho} E\left[\left(\cos\alpha-g_R\right)^2\left|Y_1,...,Y_M\right.\right]\\&+E\left[\left(\sin\alpha-g_I\right)^2\left|Y_1,...,Y_M\right.\right]+\rho\left(g_R^2+g_I^2\right)^{\frac{1}{2}}-\rho\end{aligned}, \tag{F.1}$$

which requires computation of the partial derivatives $\frac{\partial\left(E[\bullet]\right)}{\partial\rho}=0$, $\frac{\partial\left(E[\bullet]\right)}{\partial g_R}=0$, and $\frac{\partial\left(E[\bullet]\right)}{\partial g_I}=0$. The partial derivatives with respect to g_R and g_I are computed to find the solutions of $\frac{\partial\left(E[\bullet]\right)}{\partial g_R}=0$ and $\frac{\partial\left(E[\bullet]\right)}{\partial g_I}=0$ as

$$g_R\left(2+\rho\right)=2E\left[\cos\alpha\left|Y_1,...,Y_M\right.\right] \tag{F.2}$$

and

$$g_I\left(2+\rho\right)=2E\left[\sin\alpha\left|Y_1,...,Y_M\right.\right]. \tag{F.3}$$

The fundamental relationship between the real and imaginary components is given in (3.49) with

$$E\left[\cos\alpha\left|Y_1,...,Y_M\right.\right]=\frac{\int_0^{\infty}\int_0^{2\pi}\cos\alpha\, p\left(Y_1,...,Y_M\left|A,\alpha\right.\right)p\left(A,\alpha\right)d\alpha dA}{\int_0^{\infty}\int_0^{2\pi}p\left(Y_1,...,Y_M\left|A,\alpha\right.\right)p\left(A,\alpha\right)d\alpha dA} \tag{F.4}$$

and

$$E\left[\sin\alpha\left|Y_1,...,Y_M\right.\right]=\frac{\int_0^{\infty}\int_0^{2\pi}\sin\alpha\, p\left(Y_1,...,Y_M\left|A,\alpha\right.\right)p\left(A,\alpha\right)d\alpha dA}{\int_0^{\infty}\int_0^{2\pi}p\left(Y_1,...,Y_M\left|A,\alpha\right.\right)p\left(A,\alpha\right)d\alpha dA}, \tag{F.5}$$

which closely resemble the integration performed in (3.24) and (3.31) but with different arguments in the expectation operators. After substituting the statistical models for the speech prior (3.21) and noise likelihood (3.23), (F.4) and (F.5) are rewritten as

$$E\left[\cos\alpha\left|Y_1,...,Y_M\right.\right]=\frac{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_S^2}\right)\int_0^{2\pi}\cos\alpha\exp\left(-\sum_{i=1}^{M}\frac{\left|Y_i-c_iAe^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right)d\alpha dA}{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_S^2}\right)\int_0^{2\pi}\exp\left(-\sum_{i=1}^{M}\frac{\left|Y_i-c_iAe^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right)d\alpha dA} \tag{F.6}$$

and

$$E\left[\sin\alpha\left|Y_1,...,Y_M\right.\right]=\frac{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_S^2}\right)\int_0^{2\pi}\sin\alpha\exp\left(-\sum_{i=1}^{M}\frac{\left|Y_i-c_iAe^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right)d\alpha dA}{\int_0^{\infty}A\exp\left(-\frac{A^2}{\sigma_S^2}\right)\int_0^{2\pi}\exp\left(-\sum_{i=1}^{M}\frac{\left|Y_i-c_iAe^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right)d\alpha dA}. \tag{F.7}$$

By utilizing (B.2) from APPENDIX B, the inner integral over the spectral phase α in (F.6) is expanded as

$$\int_0^{2\pi}\cos\alpha\exp\left(-\sum_{i=1}^{M}\frac{\left|Y_i-c_iAe^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right)d\alpha\propto\int_0^{2\pi}\cos\alpha\exp\left(a\cos\alpha+b\sin\alpha\right)d\alpha\,. \tag{F.8}$$

Through (B.5) from APPENDIX B, the integral over the spectral phase α in (F.8) is further rewritten as

$$\int_0^{2\pi} \cos\alpha \exp\left(a\cos\alpha + b\sin\alpha\right) d\alpha = \int_0^{2\pi} \cos\alpha \cos\left(\alpha - \psi\right) d\alpha, \tag{F.9}$$

where

$$\psi = \tan^{-1}\left(\frac{b}{a}\right) \tag{F.10}$$

and a, b, and $\sqrt{a^2 + b^2}$ are given in (B.3), (B.4), and (B.6) from APPENDIX B. By using the product-to-sum cosine trigonometric identity, (F.9) simplifies to

$$\begin{aligned}\int_0^{2\pi} \cos\alpha \cos\left(\alpha - \psi\right) d\alpha &= \frac{\sqrt{a^2 + b^2}}{2}\left[\cos\psi \int_0^{2\pi} d\alpha + \int_0^{2\pi} \cos\left(2\alpha - \psi\right) d\alpha\right] \\ &= \pi\sqrt{a^2 + b^2}\cos\left(\psi\right)\end{aligned} \tag{F.11}$$

since the spectral phase shift of ψ in the second integral over the spectral phase α in (F.11) is irrelevant for the limits of integration. From (B.2) in APPENDIX B and (F.11), (F.8) is written as

$$\int_0^{2\pi} \cos\alpha \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha \propto \pi\sqrt{a^2 + b^2}\cos\left(\psi\right). \tag{F.12}$$

In a similar manner, the inner integral over the spectral phase α in (F.7) is given by

$$\int_0^{2\pi} \sin\alpha \exp\left(-\sum_{i=1}^{M} \frac{\left|Y_i - c_i A e^{j\alpha}\right|^2}{\sigma_{N_i}^2}\right) d\alpha \propto \pi\sqrt{a^2 + b^2}\cos\left(\theta\right), \tag{F.13}$$

where

$$\theta = \sin^{-1}\left(\frac{a}{\sqrt{a^2+b^2}}\right). \tag{F.14}$$

Through (F.12) and (F.13), the expectations in (F.6) and (F.7) are written as

$$E\left[\cos\alpha \middle| Y_1,...,Y_M\right] = \frac{\sqrt{a^2+b^2}}{2}\cos(\psi)\frac{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)dA}{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\sum_{i=1}^{M}\frac{c_iY_i}{\sigma_{N_i}^2}\right|\right)dA} \tag{F.15}$$

and

$$E\left[\sin\alpha \middle| Y_1,...,Y_M\right] = \frac{\sqrt{a^2+b^2}}{2}\cos(\theta)\frac{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)dA}{\int_0^\infty A\exp\left(-A^2\frac{1}{\lambda}\right)I_0\left(2A\left|\sum_{i=1}^{M}\frac{c_iY_i}{\sigma_{N_i}^2}\right|\right)dA} \tag{F.16}$$

with $\frac{1}{\lambda}$ given by (3.26). By utilizing the expectations from (F.15) and (F.16) and employing the definitions (F.10) and (F.14), the closed-form solution of $\hat{\alpha}$ from (3.50) is written as

$$\begin{aligned}\hat{\alpha} &= \tan^{-1}\left(\frac{\cos\theta}{\cos\psi}\right)\\ &= \tan^{-1}\left(\frac{b}{a}\right)\end{aligned}, \tag{F.17}$$

where a and b are specified in (B.3) and (B.4) from APPENDIX B. Through simplification of the ratio $\frac{b}{a}$ using $A_i = c_i A$ and $\sigma_{S_i}^2 = c_i^2\sigma_S^2$, the final closed-form solution of $\hat{\alpha}$ in (F.17) is given in (3.50) as

$$\hat{\alpha} = \tan^{-1}\left(\frac{\sum_{i=1}^{M}\frac{\sqrt{\xi_i}}{\sigma_{N_i}}\mathrm{Im}\left(Y_i\right)}{\sum_{i=1}^{M}\frac{\sqrt{\xi_i}}{\sigma_{N_i}}\mathrm{Re}\left(Y_i\right)}\right). \tag{F.18}$$

APPENDIX G COMPLEX RE/IM SPECTRAL COMP ESTIMATOR

In this appendix, the minimum mean-square error complex real and imaginary spectral component estimator is derived for distributed multichannel signals using Gaussian noise and Gaussian speech statistical models. After substitution of (3.51) and (3.53), (3.54) is written as

$$\hat{S}_{R,I} = \frac{\int_{-\infty}^{\infty} S_{R,I} \exp\left(-\left[\sum_{i=1}^{M} \frac{\left(Y_{i,(R,I)} - c_i S_{R,I} \right)^2}{\sigma_{N_i}^2} + \frac{S_{R,I}}{\sigma_S^2} \right] \right) dS_{R,I}}{\int_{-\infty}^{\infty} \exp\left(-\left[\sum_{i=1}^{M} \frac{\left(Y_{i,(R,I)} - c_i S_{R,I} \right)^2}{\sigma_{N_i}^2} + \frac{S_{R,I}}{\sigma_S^2} \right] \right) dS_{R,I}} \tag{G.1}$$

or

$$\hat{S}_{R,I} = \frac{\int_{-\infty}^{\infty} S_{R,I} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I} \left(\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2} \right) \right) dS_{R,I}}{\int_{-\infty}^{\infty} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I} \left(\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2} \right) \right) dS_{R,I}}, \tag{G.2}$$

where

$$\frac{1}{\lambda} = \frac{1}{\sigma_S^2} + \sum_{i=1}^{M} \frac{c_i^2}{\sigma_{N_i}^2}. \tag{G.3}$$

By splitting the integral in both the numerator and denominator in (G.2) each into two separate integrals and utilizing the relationship 3.462.1 in [25],

$$\begin{aligned}&\int_{-\infty}^{\infty} S_{R,I}\exp\left(-S_{R,I}^2\frac{1}{\lambda}+2S_{R,I}\left(\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}\right)\right)dS_{R,I}\\&=\left(2\frac{1}{\lambda}\right)^{-1}\exp\left(\frac{\left(\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}\right)^2}{2\frac{1}{\lambda}}\right)\left[D_{-2}\left(\frac{-\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)-D_{-2}\left(\frac{\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)\right]\end{aligned}\tag{G.4}$$

and

$$\begin{aligned}&\int_{-\infty}^{\infty}\exp\left(-S_{R,I}^2\frac{1}{\lambda}+2S_{R,I}\left(\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}\right)\right)dS_{R,I}\\&=\left(2\frac{1}{\lambda}\right)^{-\frac{1}{2}}\exp\left(\frac{\left(\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}\right)^2}{2\frac{1}{\lambda}}\right)\left[D_{-1}\left(\frac{\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)+D_{-1}\left(\frac{-\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)\right],\end{aligned}\tag{G.5}$$

where $D_{\bullet}(\bullet)$ is the parabolic cylinder function defined by 9.240 in [25]. With (G.4) and (G.5), (G.2) is rewritten as

$$\hat{S}_{R,I}=\left(2\frac{1}{\lambda}\right)^{-\frac{1}{2}}\left[\frac{D_{-2}\left(-\frac{\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)-D_{-2}\left(\frac{\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)}{D_{-1}\left(\frac{\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)+D_{-1}\left(-\frac{\sqrt{2}\sum_{i=1}^{M}\frac{c_iY_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)}\right],\tag{G.6}$$

where

$$\left(2\frac{1}{\lambda}\right)^{-\frac{1}{2}}=\frac{\sqrt{2}}{2}\left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i}\right)^{\frac{1}{2}}\tag{G.7}$$

with $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$. The arguments to the parabolic cylinder functions are simplified to

$$\frac{\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} = \frac{\sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}}}{\left(1+\sum_{i=1}^{M} \xi_i\right)^{\frac{1}{2}}} \tag{G.8}$$

with $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$ and defined as

$$N_{(R,I)^{\pm}} = \sqrt{2} \frac{\sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} Y_{i,(R,I)}}{\left(1+\sum_{i=1}^{M} \xi_i\right)^{\frac{1}{2}}} \tag{G.9}$$

using the same notation as in [13]. Through the substitution of (G.7) and (G.9), (G.6) is rewritten as

$$\hat{S}_{R,I} = \frac{\sqrt{2}}{2} \left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M} \xi_i} \right)^{\frac{1}{2}} \left[\frac{D_{-2}\left(N_{(R,I)-}\right) - D_{-2}\left(N_{(R,I)+}\right)}{D_{-1}\left(N_{(R,I)+}\right) + D_{-1}\left(N_{(R,I)-}\right)} \right]. \tag{G.10}$$

By simplifying the ratio of parabolic cylinder functions in (G.10), the ratio is now given as

$$\frac{D_{-2}\left(-\frac{\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \right) - D_{-2}\left(\frac{\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \right)}{D_{-1}\left(\frac{\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \right) + D_{-1}\left(-\frac{\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} \right)} = \sqrt{2}\left(\frac{\sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} Y_{i,(R,I)}}{\left(1+\sum_{i=1}^{M} \xi_i\right)^{\frac{1}{2}}} \right) = N_{(R,I)+} . \tag{G.11}$$

Thus, the final closed-form solution $\hat{S}_{R,I}$ is given as in (3.55) as

$$\hat{S}_{R,I} = \frac{\sigma_S \sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} Y_{i,(R,I)}}{1 + \sum_{i=1}^{M} \xi_i}. \qquad \text{(G.12)}$$

APPENDIX H SUPPLEMENTARY EXPERIMENTAL RESULTS

In this appendix, supplementary experimental results are presented using primarily the SNR metric but also the SSNR metric for four basic sets of experiments: enhancement, spectral phase estimation, time alignment, and attenuation factor estimation.

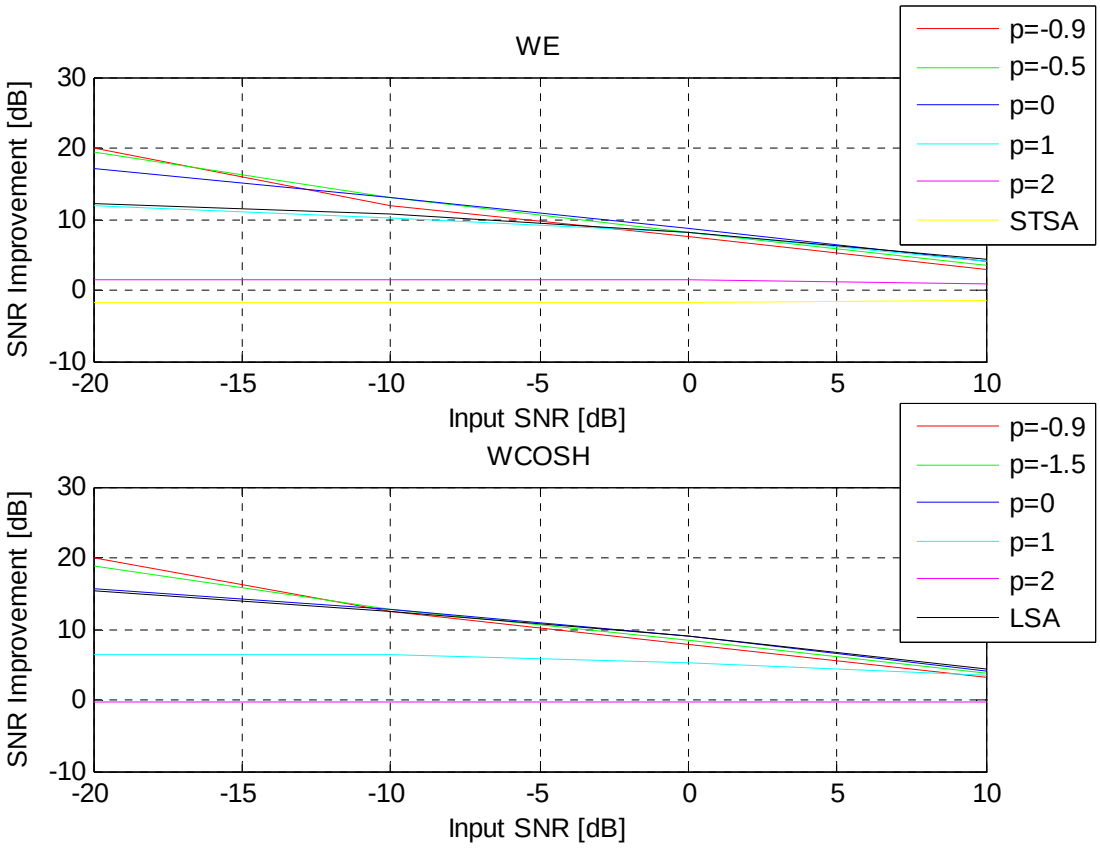

Figure H-1 SNR Improvements for Single Channel Weighted Euclidean (WE) and Single Channel Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Single Channel Spectral Phase Estimation

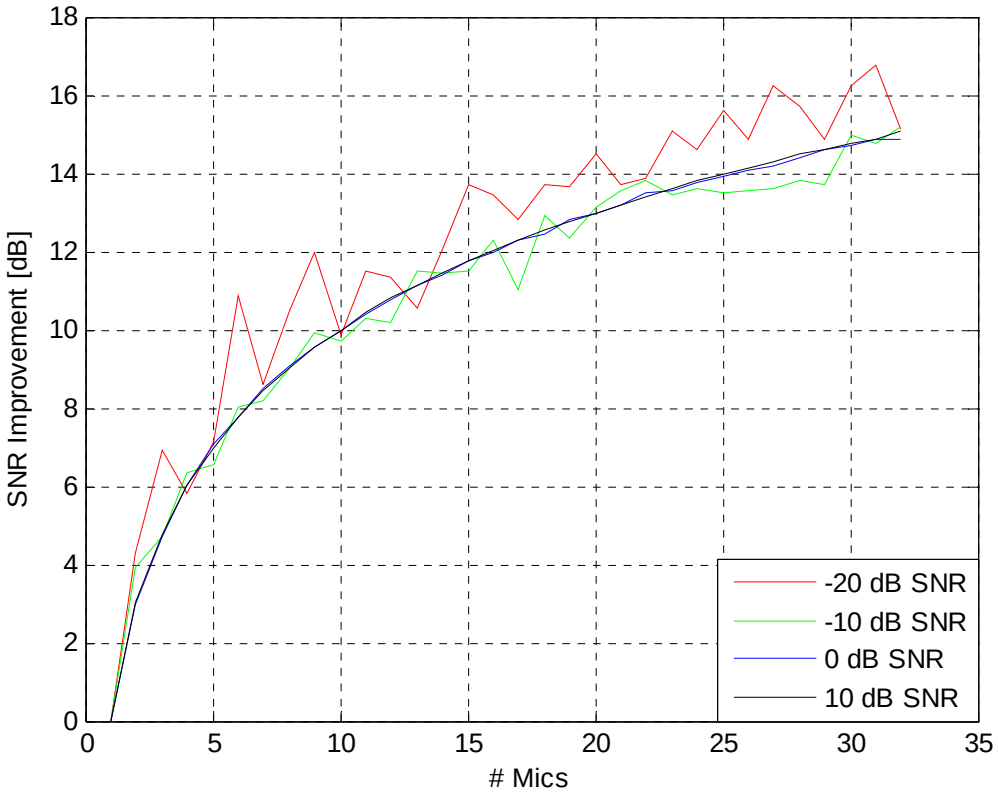

Figure H-2 SNR Improvements for Time Domain Estimation (Unity Attenuation Factors)

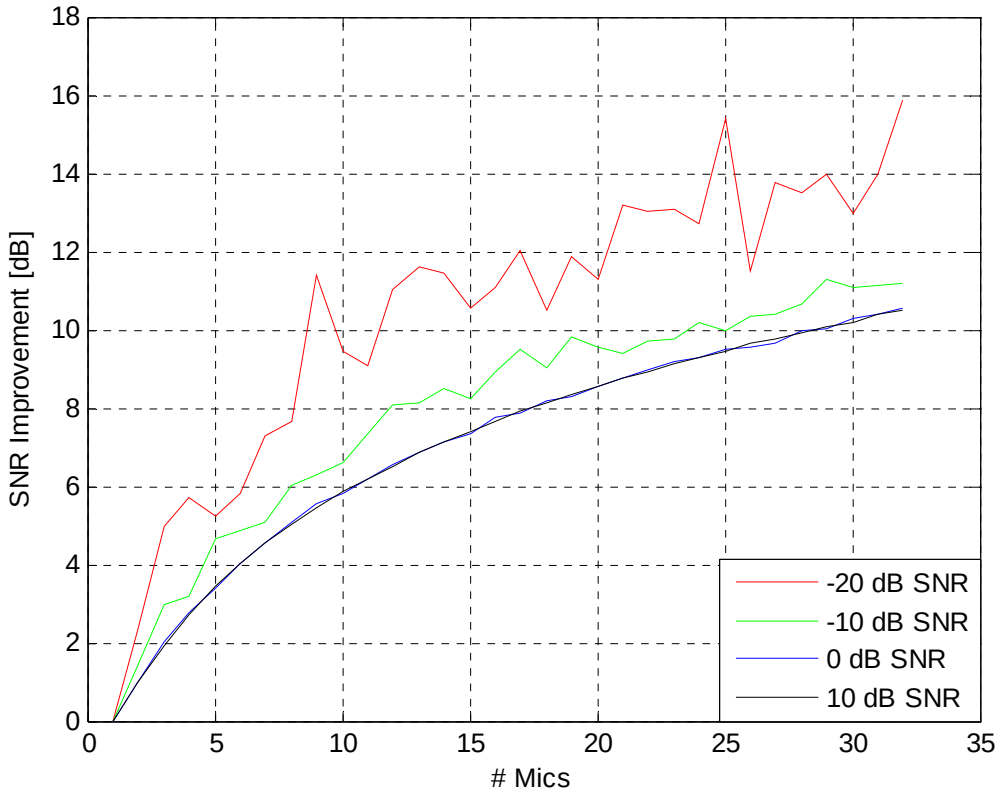

Figure H-3 SNR Improvements for Time Domain Estimation (Linear Attenuation Factors)

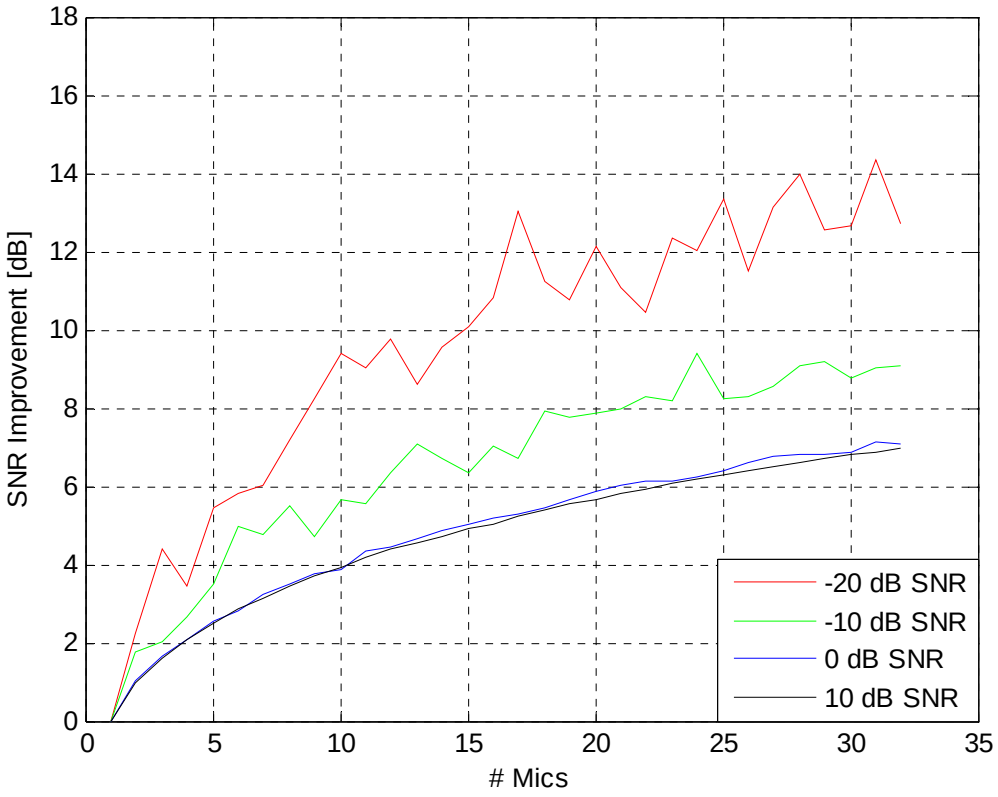

Figure H-4 SNR Improvements for Time Domain Estimation (Logarithmic Attenuation Factors)

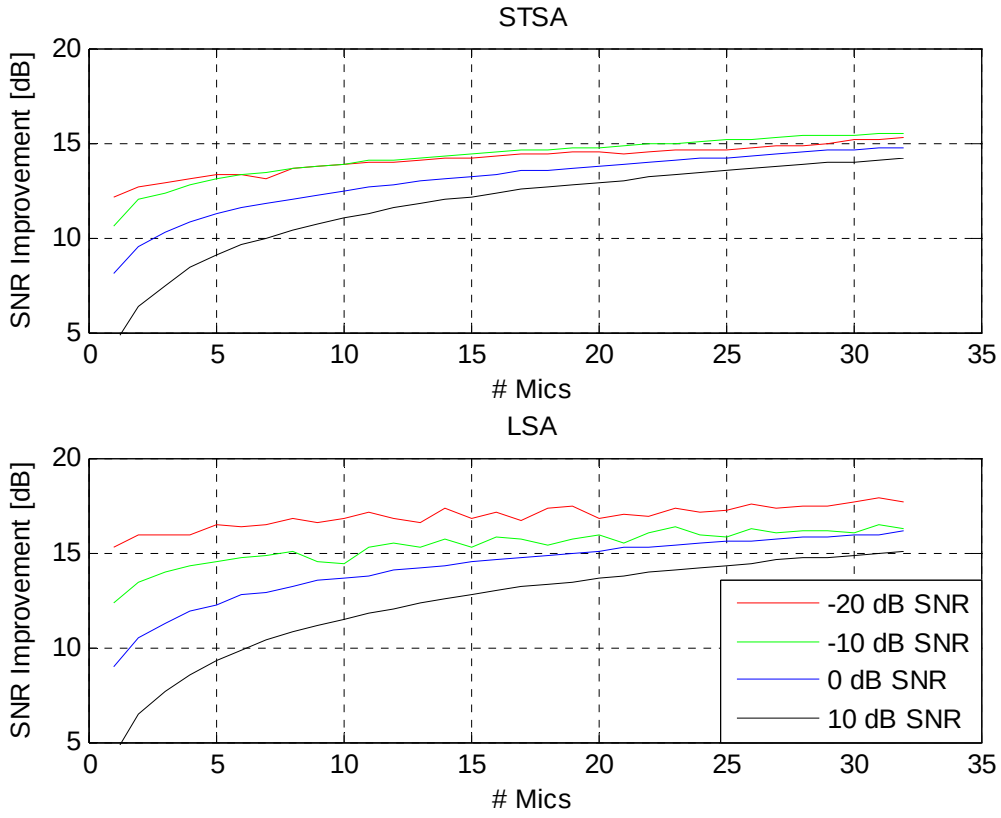

Figure H-5 SNR Improvements for Short-Time Spectral Amplitude (STSA) and Log-Spectral Amplitude (LSA) Estimation with Spectral Phase Estimation (Unity Attenuation Factors)

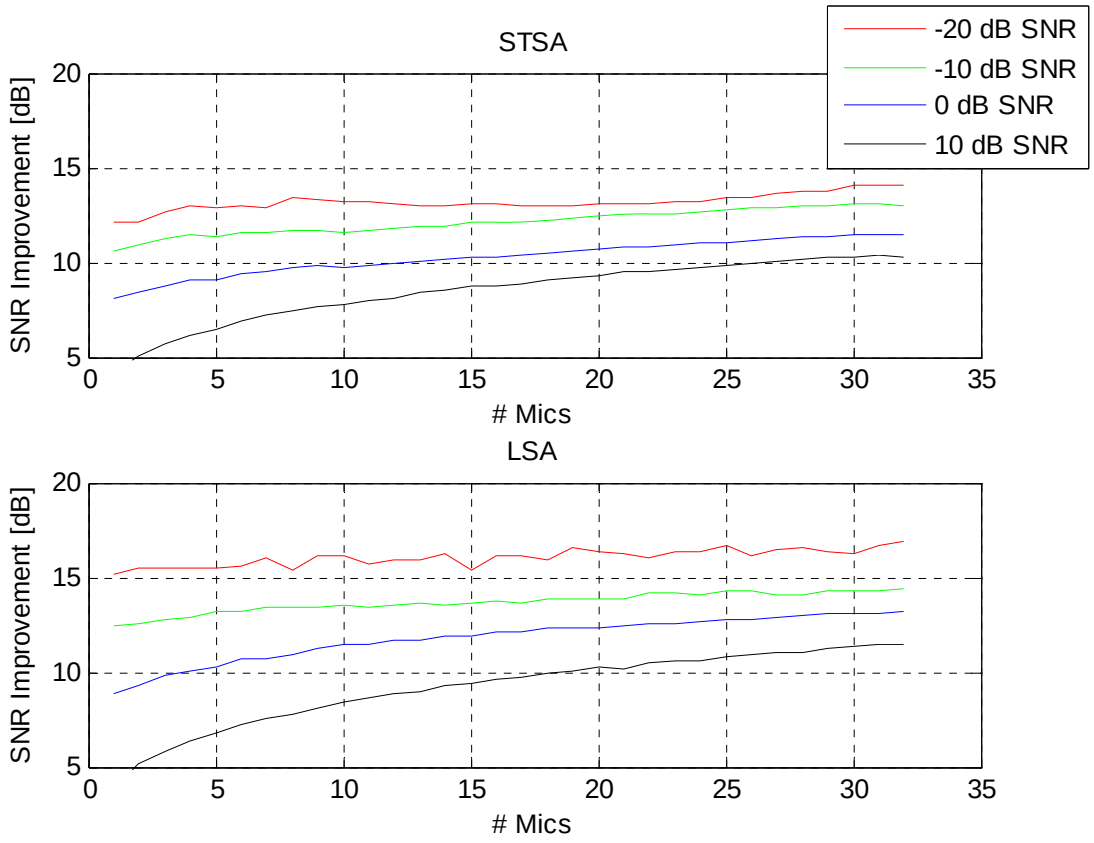

Figure H-6 SNR Improvements for Short-Time Spectral Amplitude (STSA) and Log-Spectral Amplitude (LSA) Estimation with Spectral Phase Estimation (Linear Attenuation Factors)

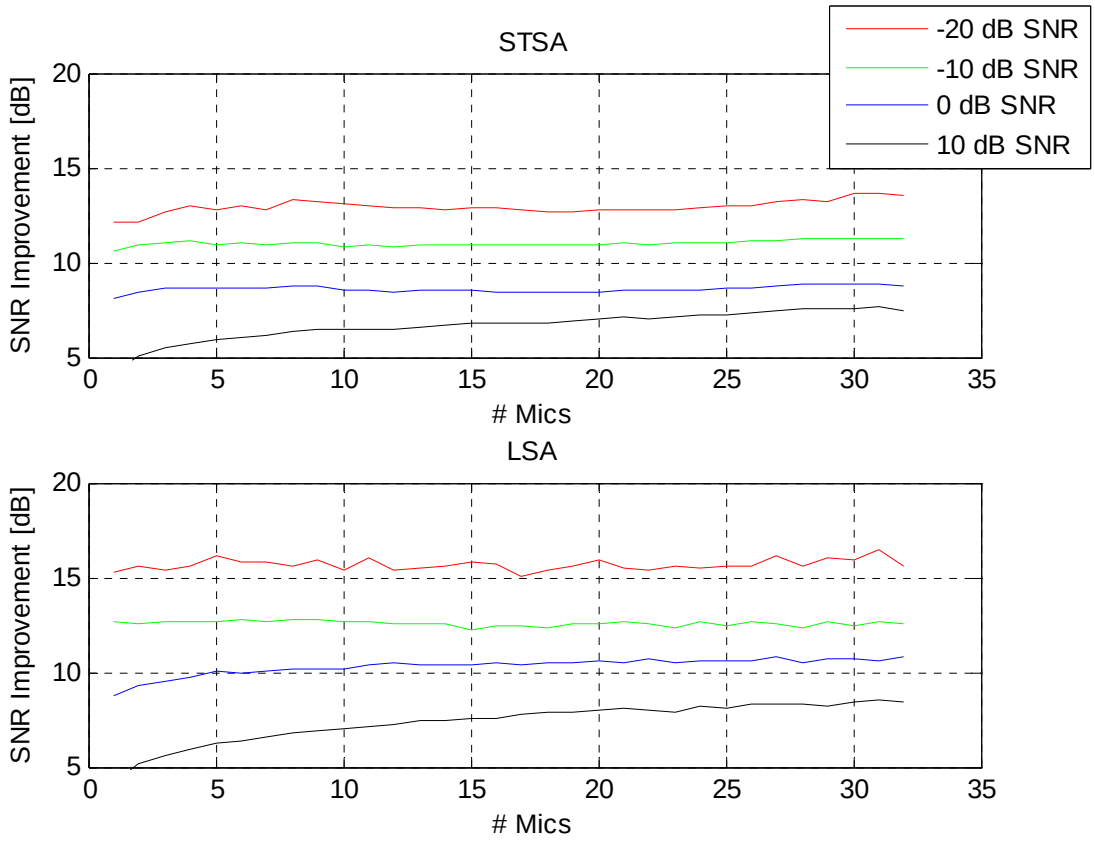

Figure H-7 SNR Improvements for Short-Time Spectral Amplitude (STSA) and Log-Spectral Amplitude (LSA) Estimation with Spectral Phase Estimation (Logarithmic Attenuation Factors)

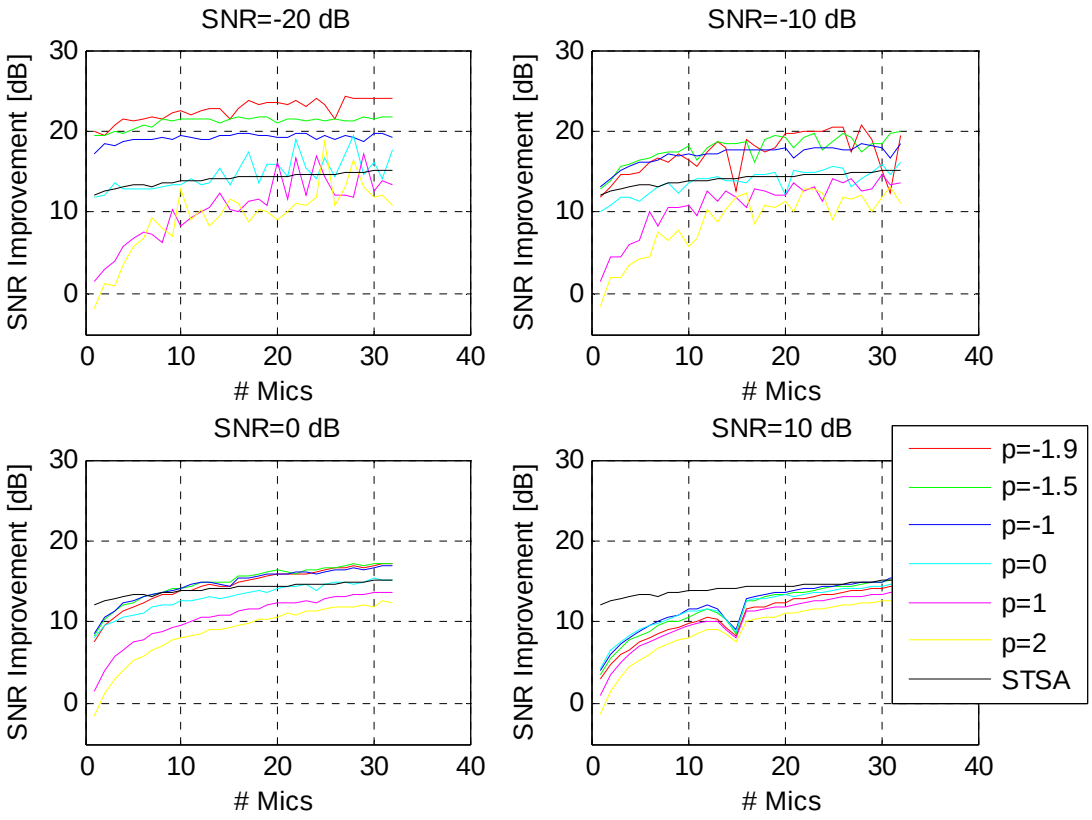

Figure H-8 SNR Improvements for Weighted Euclidean (WE) Spectral Amplitude Estimation with Spectral Phase Estimation (Unity Attenuation Factors)

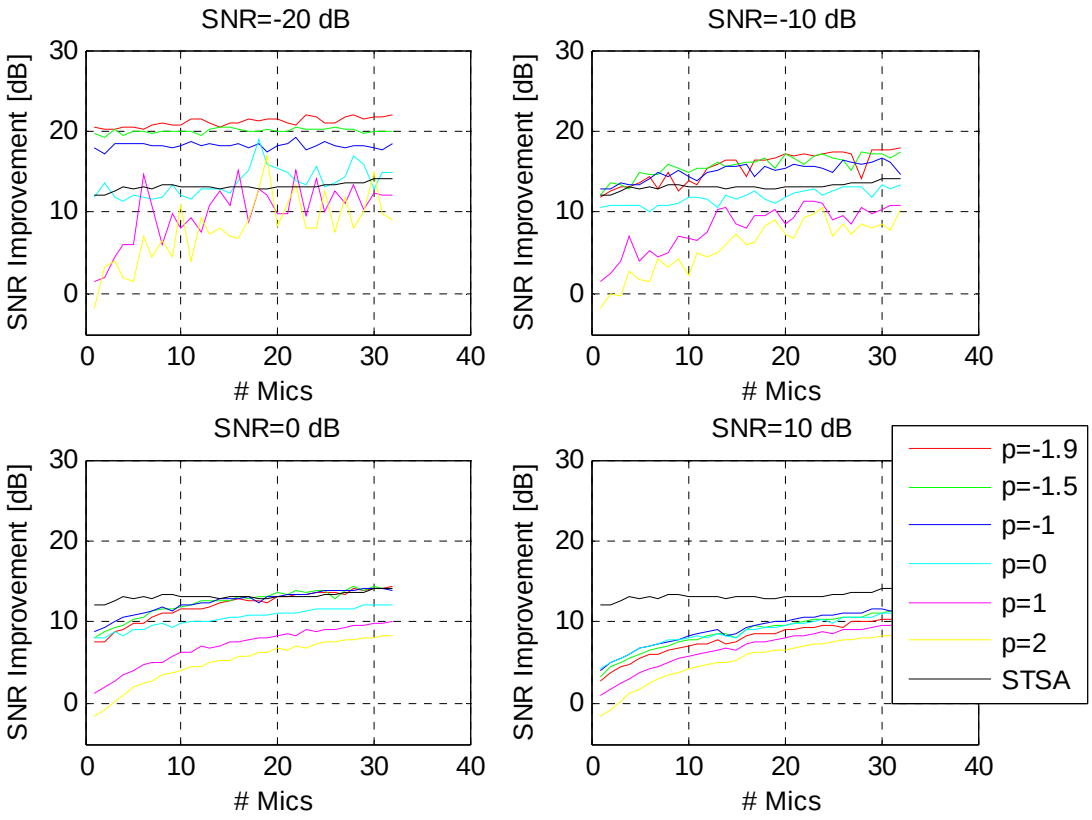

Figure H-9 SNR Improvements for Weighted Euclidean (WE) Spectral Amplitude Estimation with Spectral Phase Estimation (Linear Attenuation Factors)

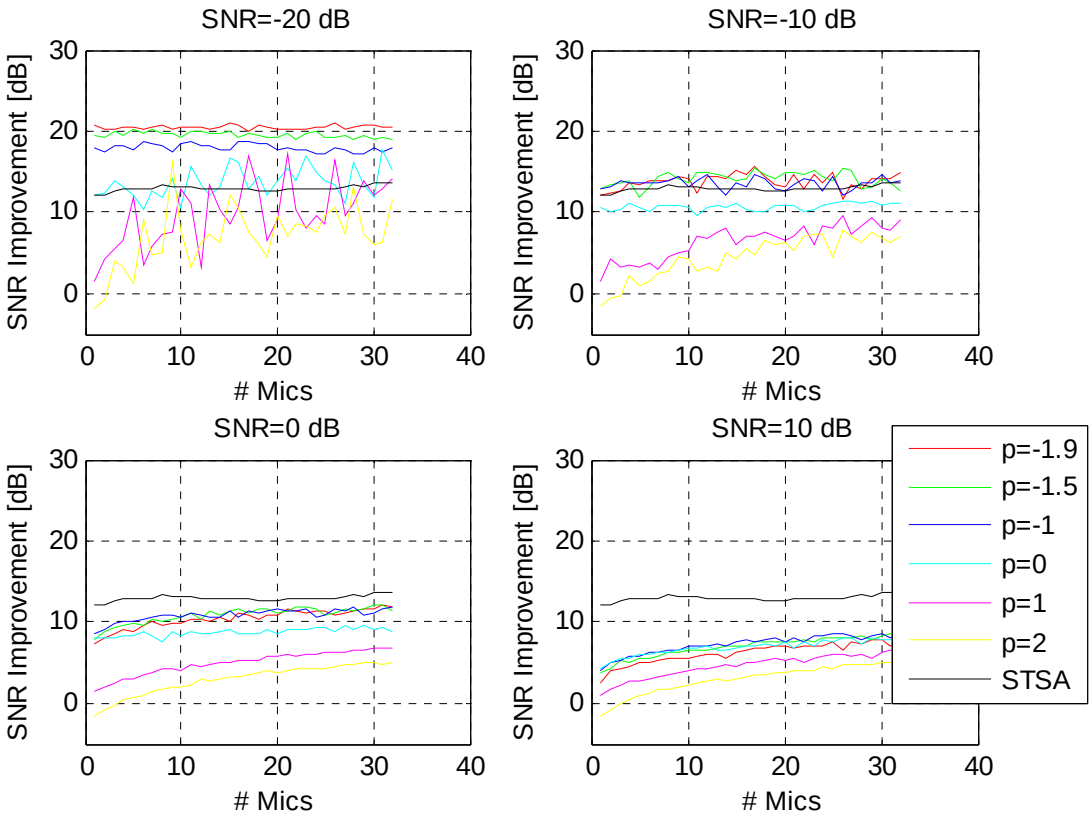

Figure H-10 SNR Improvements for Weighted Euclidean (WE) Spectral Amplitude Estimation with Spectral Phase Estimation (Logarithmic Attenuation Factors)

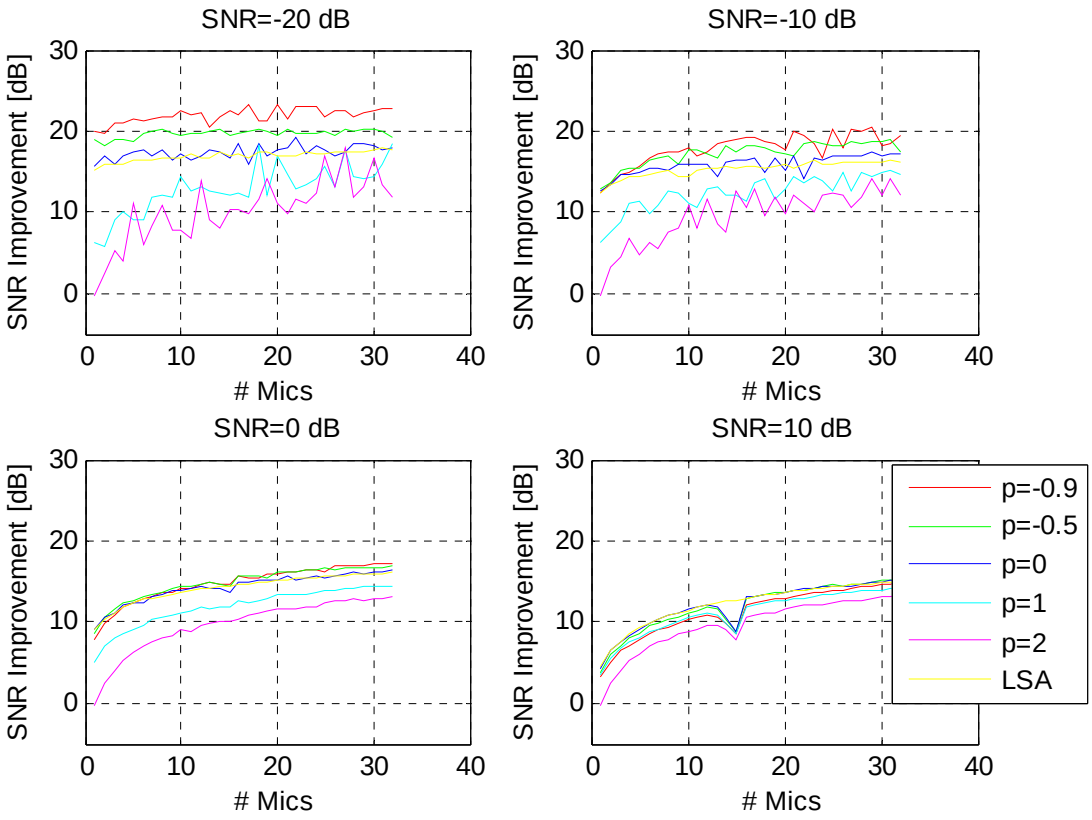

Figure H-11 SNR Improvements for Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Spectral Phase Estimation (Unity Attenuation Factors)

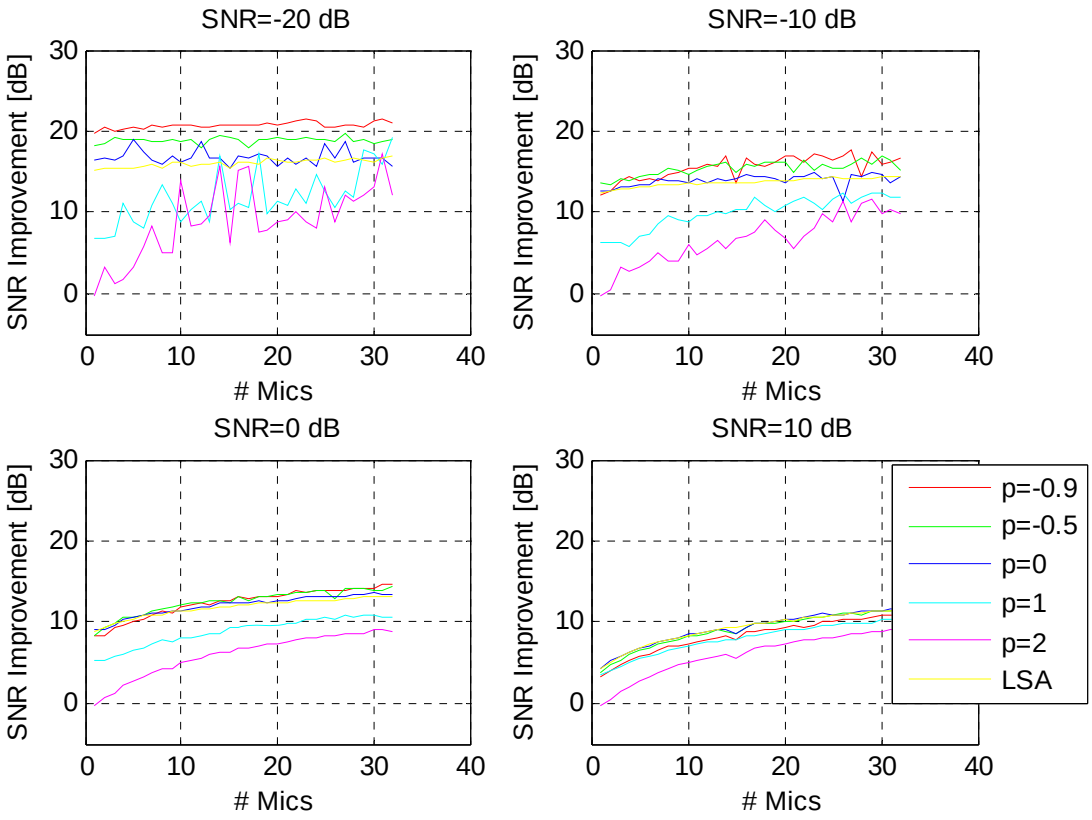

Figure H-12 SNR Improvements for Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Spectral Phase Estimation (Linear Attenuation Factors)

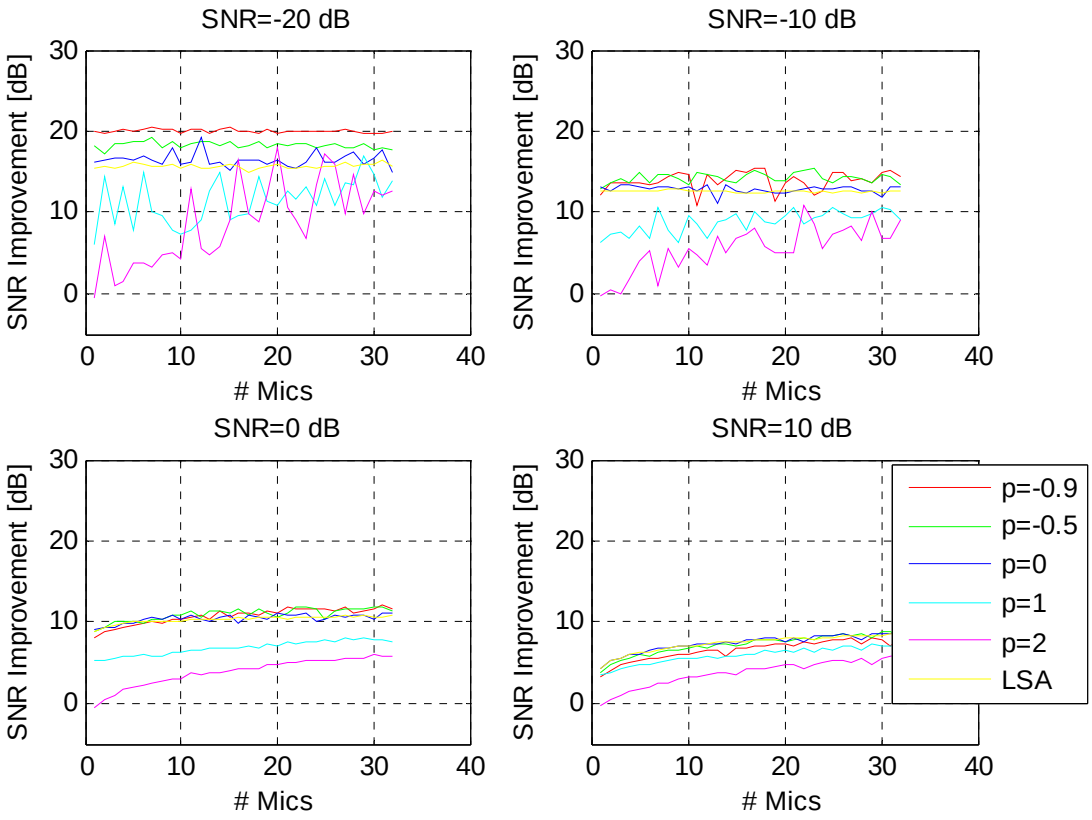

Figure H-13 SNR Improvements for Weighted Cosh (WCOSH) Spectral Amplitude Estimation with Spectral Phase Estimation (Logarithmic Attenuation Factors)

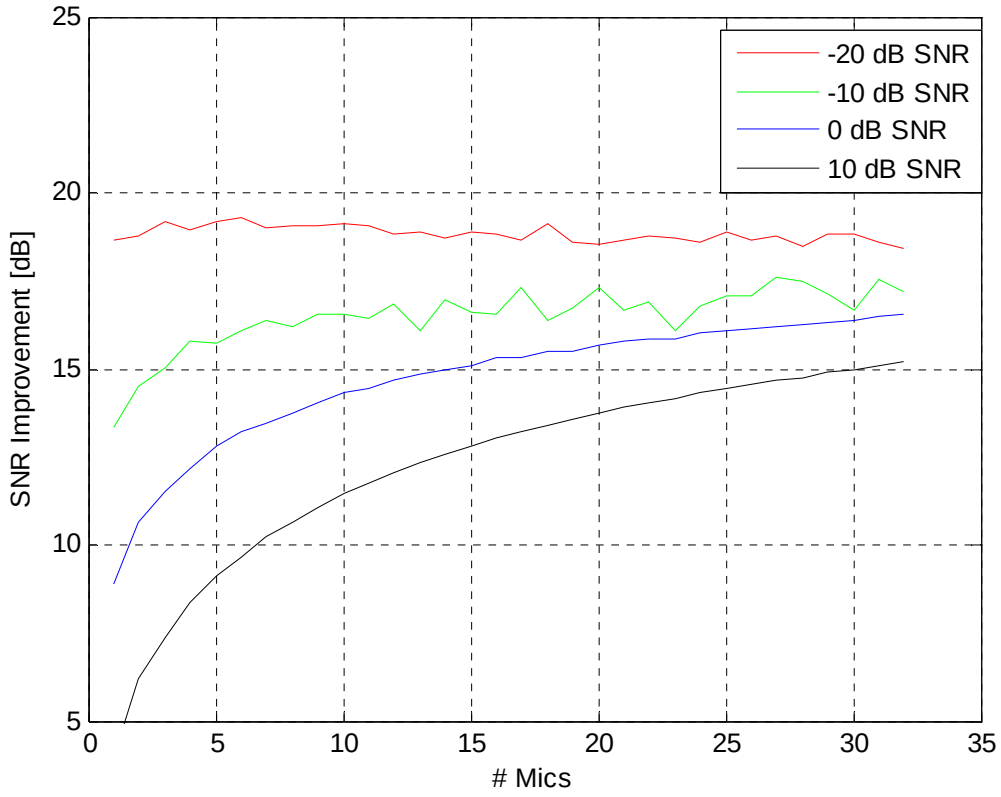

Figure H-14 SNR Improvements for Complex Real and Imaginary Spectral Component Estimation (Unity Attenuation Factors)

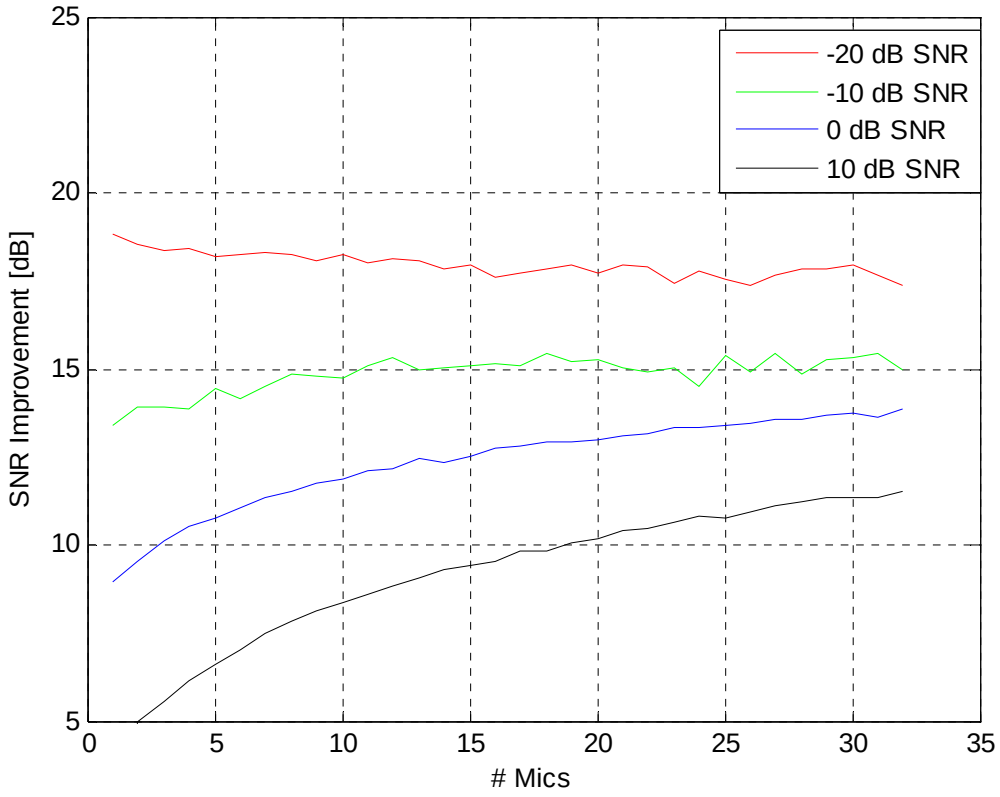

Figure H-15 SNR Improvements for Complex Real and Imaginary Spectral Component Estimation (Linear Attenuation Factors)

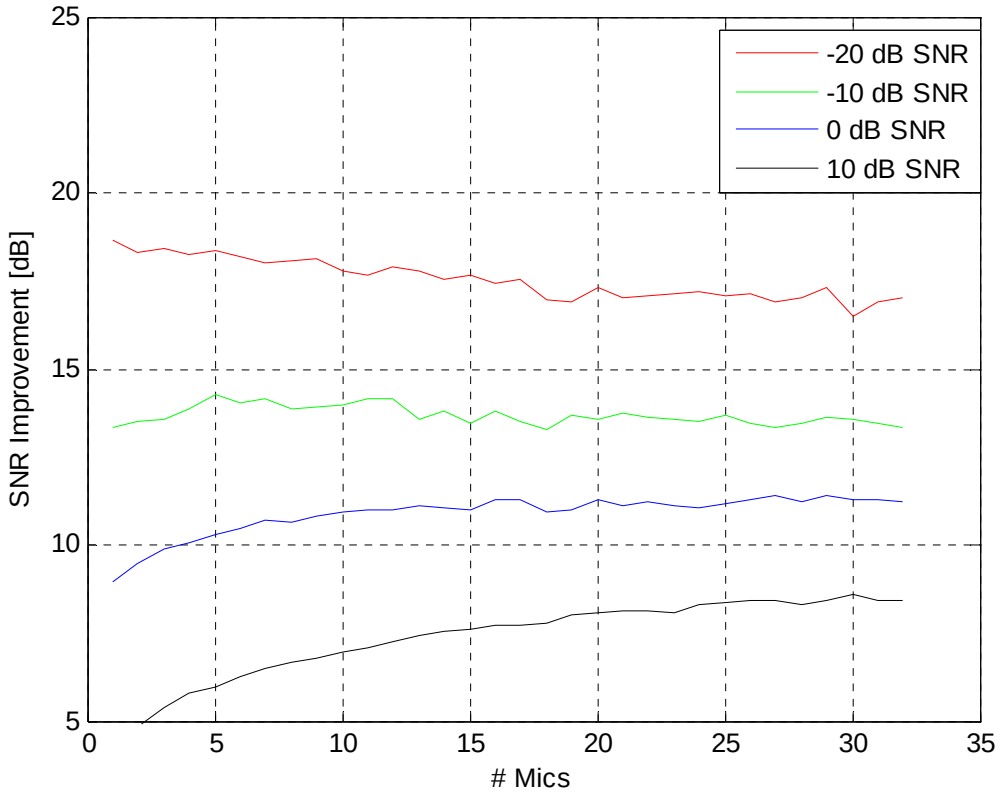

Figure H-16 SNR Improvements for Complex Real and Imaginary Spectral Component Estimation (Logarithmic Attenuation Factors)

SNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	15.12	0.00	15.89	0.00	12.73
STSA + Estimated Phase	12.07	15.22	12.07	14.10	12.07	13.59
STSA + Noisy Phase	12.07	14.32	12.07	13.71	12.07	13.42
LSA + Estimated Phase	15.29	17.62	15.20	16.93	15.31	15.63
LSA + Noisy Phase	15.26	16.48	15.61	16.40	15.26	15.38
WE + Estimated Phase (p=-1.9)	**20.02**	**23.92**	**20.43**	**21.96**	**20.68**	**20.43**
WCOSH + Estimated Phase (p=-0.9)	20.09	22.87	19.75	20.98	20.00	19.86
Complex	18.63	18.41	18.81	17.36	18.68	17.02

Table H-1 SNR Improvement (Input SNR/SSNR = -20.0 dB/-27.6 dB)

SNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	15.20	0.00	11.21	0.00	9.07
STSA + Estimated Phase	10.62	15.45	10.62	13.03	10.62	11.21
STSA + Noisy Phase	10.62	13.52	10.62	12.19	10.62	10.94
LSA + Estimated Phase	12.32	16.23	12.45	14.43	12.68	12.57
LSA + Noisy Phase	12.59	13.69	12.47	12.95	12.40	12.16
WE + Estimated Phase (p=-1.9)	**11.92**	**13.37**	**11.88**	**18.05**	**12.21**	**14.80**
WCOSH + Estimated Phase (p=-0.9)	12.35	19.49	12.09	16.61	12.04	14.53
Complex	13.35	17.19	13.41	14.98	13.35	13.35

Table H-2 SNR Improvement (Input SNR/SSNR = -10.0 dB/-17.6 dB)

SNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	14.89	0.00	10.57	0.00	7.07
STSA + Estimated Phase	8.10	14.76	8.10	11.42	8.10	8.72
STSA + Noisy Phase	8.10	10.51	8.10	9.38	8.10	7.93
LSA + Estimated Phase	8.97	16.09	8.80	13.22	8.73	10.84
LSA + Noisy Phase	8.83	10.93	8.87	10.49	8.87	9.46
WE + Estimated Phase (p=-1.9)	**7.50**	**17.23**	**7.47**	**14.30**	**7.40**	**11.92**
WCOSH + Estimated Phase (p=-0.9)	7.77	17.27	8.26	14.56	7.95	11.71
Complex	8.89	16.54	8.92	13.85	8.96	11.25

Table H-3 SNR Improvement (Input SNR/SSNR = 0.0 dB/-7.6 dB)

SNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	15.05	0.00	10.49	0.00	6.99
STSA + Estimated Phase	4.37	14.17	4.37	10.28	4.37	7.46
STSA + Noisy Phase	4.37	7.72	4.37	7.01	4.37	6.03
LSA + Estimated Phase	4.39	15.02	4.31	11.49	4.38	8.38
LSA + Noisy Phase	4.39	7.93	4.39	7.65	4.38	6.88
WE + Estimated Phase (p=-1.9)	**2.97**	**14.48**	**2.80**	**10.51**	**2.47**	**7.86**
WCOSH + Estimated Phase (p=-0.9)	3.31	14.86	3.22	10.97	3.12	8.03
Complex	4.11	15.19	4.11	11.50	4.07	8.44

Table H-4 SNR Improvement (Input SNR/SSNR = 10.0 dB/2.4 dB)

SSNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	15.12	0.00	16.88	0.00	12.83
STSA + Estimated Phase	15.13	15.37	12.49	14.26	12.49	13.79
STSA + Noisy Phase	12.49	14.57	12.49	13.92	12.49	13.65
LSA + Estimated Phase	16.02	17.84	15.88	17.45	16.02	15.96
LSA + Noisy Phase	15.92	16.85	16.35	16.90	15.94	15.74
WE + Estimated Phase (p=-1.9)	**25.72**	**25.41**	**24.91**	**23.55**	**25.45**	**23.45**
WCOSH + Estimated Phase (p=-0.9)	23.67	23.42	23.39	22.26	23.59	21.95
Complex	20.59	18.63	20.86	17.65	20.64	17.59

Table H-5 SSNR Improvement (Input SNR/SSNR = -20.0 dB/-27.6 dB)

SSNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	15.62	0.00	11.27	0.00	9.25
STSA + Estimated Phase	15.41	16.05	11.17	13.76	11.17	12.16
STSA + Noisy Phase	11.17	14.35	11.17	13.00	11.17	11.90
LSA + Estimated Phase	13.34	16.57	13.54	14.80	13.79	13.34
LSA + Noisy Phase	13.64	14.50	13.60	13.74	13.53	13.12
WE + Estimated Phase (p=-1.9)	**16.88**	**21.38**	**16.78**	**19.93**	**16.96**	**17.96**
WCOSH + Estimated Phase (p=-0.9)	16.70	20.63	16.18	18.54	16.36	17.16
Complex	15.90	17.55	15.97	15.94	15.90	14.62

Table H-6 SSNR Improvement (Input SNR/SSNR = -10.0 dB/-17.6 dB)

SSNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	14.89	0.00	10.57	0.00	7.08
STSA + Estimated Phase	8.86	14.94	8.86	11.67	8.86	9.20
STSA + Noisy Phase	8.86	11.20	8.86	9.87	8.86	8.47
LSA + Estimated Phase	10.42	16.35	10.21	13.59	10.13	11.49
LSA + Noisy Phase	10.28	11.83	10.25	11.21	10.28	10.45
WE + Estimated Phase (p=-1.9)	**11.27**	**18.73**	**11.31**	**16.36**	**11.14**	**14.38**
WCOSH + Estimated Phase (p=-0.9)	10.97	18.44	11.29	16.15	10.91	14.09
Complex	11.35	16.90	11.15	14.45	11.30	12.37

Table H-7 SSNR Improvement (Input SNR/SSNR = 0.0 dB/-7.6 dB)

SSNR Improvements	**Attenuation Factors**					
Methods	**Unity**		**Linear**		**Logarithmic**	
	1-Ch.	**32-Ch.**	**1-Ch.**	**32-Ch.**	**1-Ch.**	**32-Ch.**
Time Domain	0.00	15.05	0.00	10.48	0.00	6.99
STSA + Estimated Phase	2.40	14.01	14.38	5.64	10.62	5.64
STSA + Noisy Phase	5.64	8.62	5.64	7.68	5.64	6.59
LSA + Estimated Phase	6.24	15.35	6.26	12.00	6.25	9.44
LSA + Noisy Phase	6.25	9.03	6.27	8.60	6.27	7.86
WE + Estimated Phase (p=-1.9)	**6.25**	**15.93**	**6.10**	**12.42**	**5.63**	**10.11**
WCOSH + Estimated Phase (p=-0.9)	6.37	16.08	6.27	12.65	6.06	10.17
Complex	6.67	15.66	6.67	12.35	6.60	9.80

Table H-8 SSNR Improvement (Input SNR/SSNR = 10.0 dB/2.4 dB)

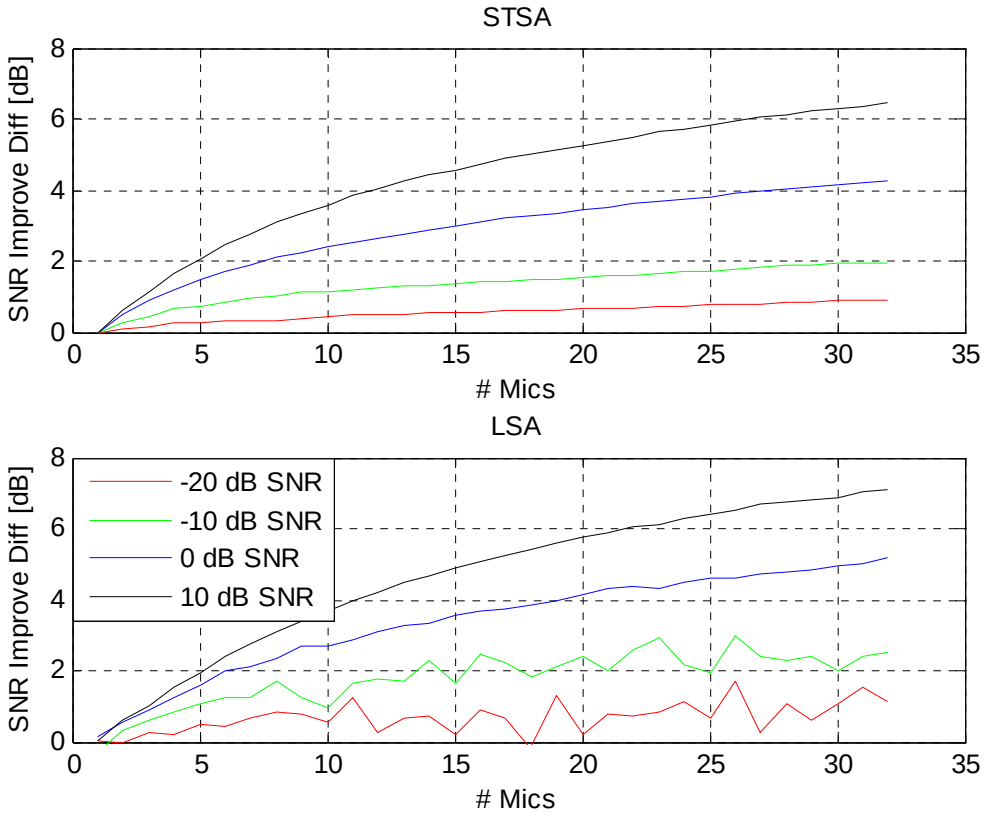

Figure H-17 SNR Improvement Difference between Multichannel Short-Time Spectral Amplitude (STSA) and Multichannel Log-Spectral Amplitude (LSA) Estimation with Multichannel Spectral Phase Estimation and Single Channel Short-Time Spectral Amplitude (STSA) and Single Channel Log-Spectral Amplitude (LSA) Estimation with Single Channel (Noisy) Spectral Phase Estimation (Unity Attenuation Factors)

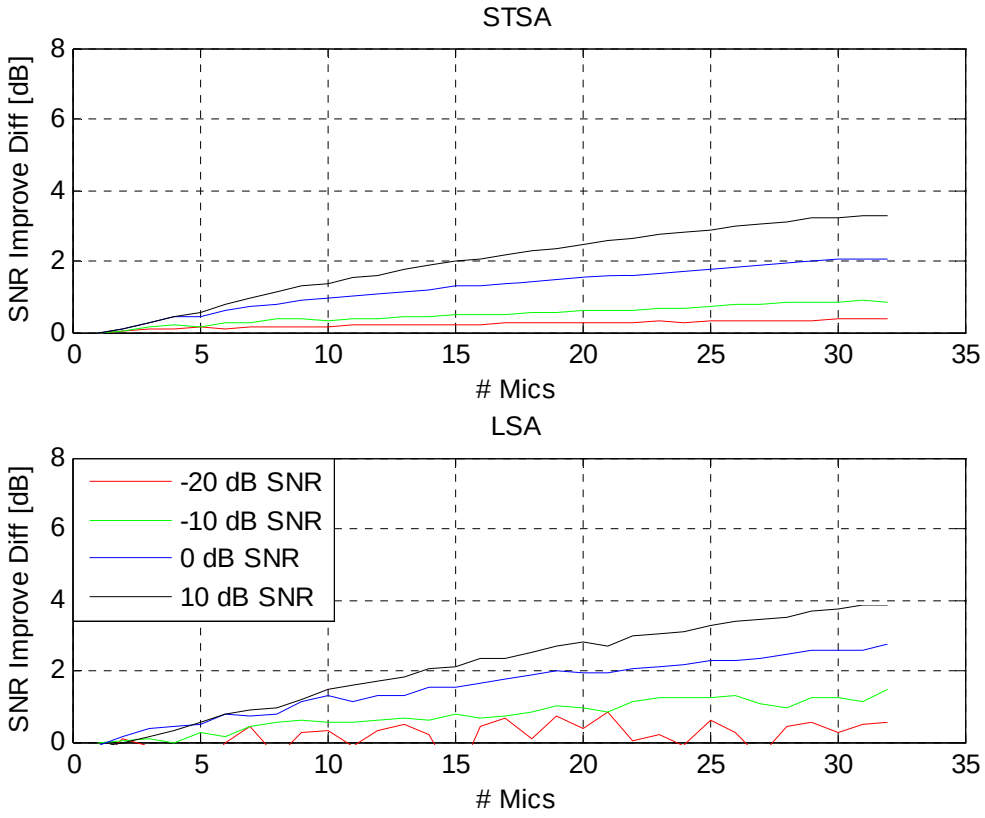

Figure H-18 SNR Improvement Difference between Multichannel Short-Time Spectral Amplitude (STSA) and Multichannel Log-Spectral Amplitude (LSA) Estimation with Multichannel Spectral Phase Estimation and Single Channel Short-Time Spectral Amplitude (STSA) and Single Channel Log-Spectral Amplitude (LSA) Estimation with Single Channel (Noisy) Spectral Phase Estimation (Linear Attenuation Factors)

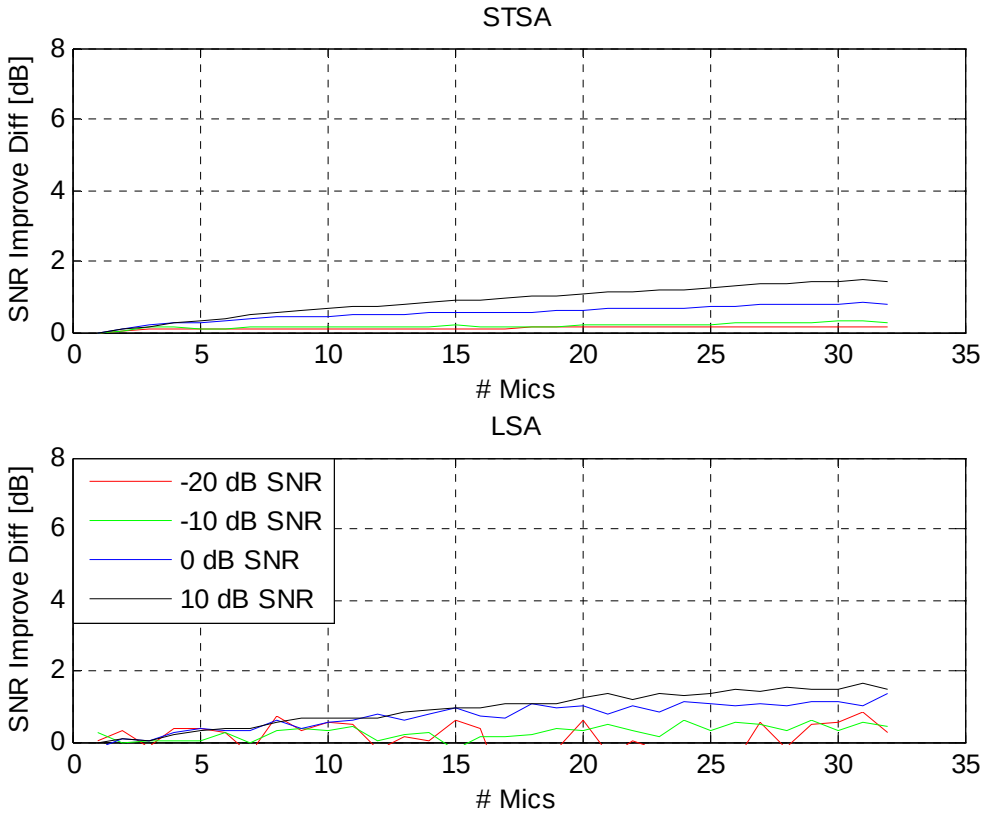

Figure H-19 SNR Improvement Difference between Multichannel Short-Time Spectral Amplitude (STSA) and Multichannel Log-Spectral Amplitude (LSA) Estimation with Multichannel Spectral Phase Estimation and Single Channel Short-Time Spectral Amplitude (STSA) and Single Channel Log-Spectral Amplitude (LSA) Estimation with Single Channel (Noisy) Spectral Phase Estimation (Logarithmic Attenuation Factors)

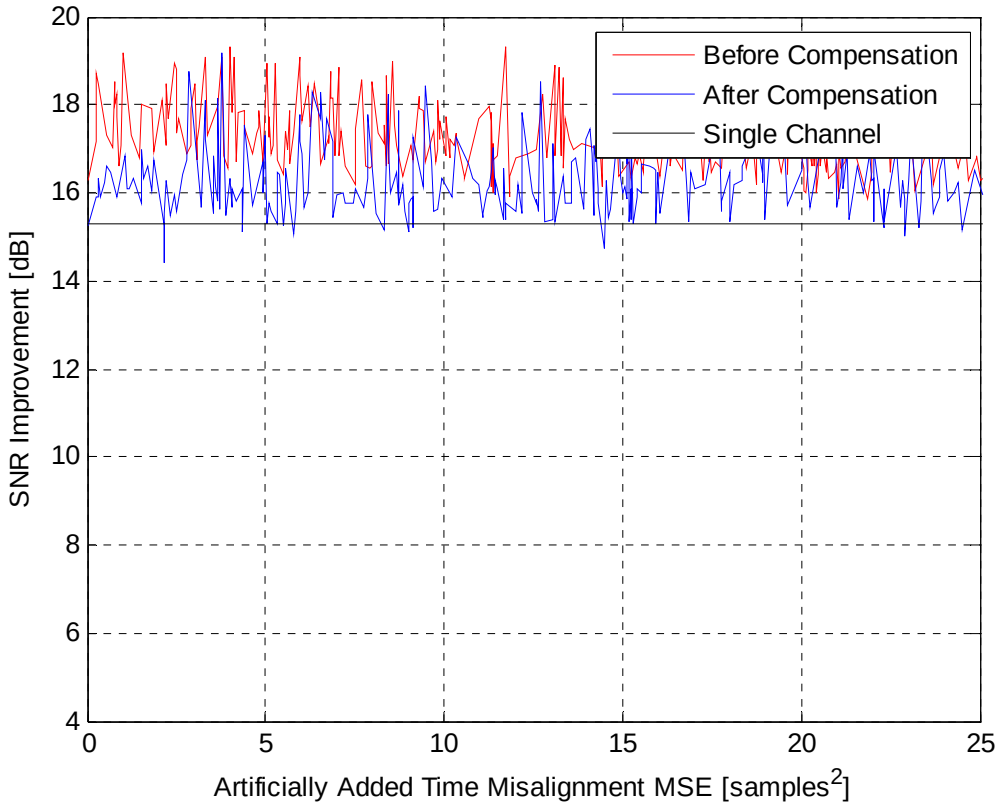

Figure H-20 SNR Improvement for 32 Microphones, -20 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

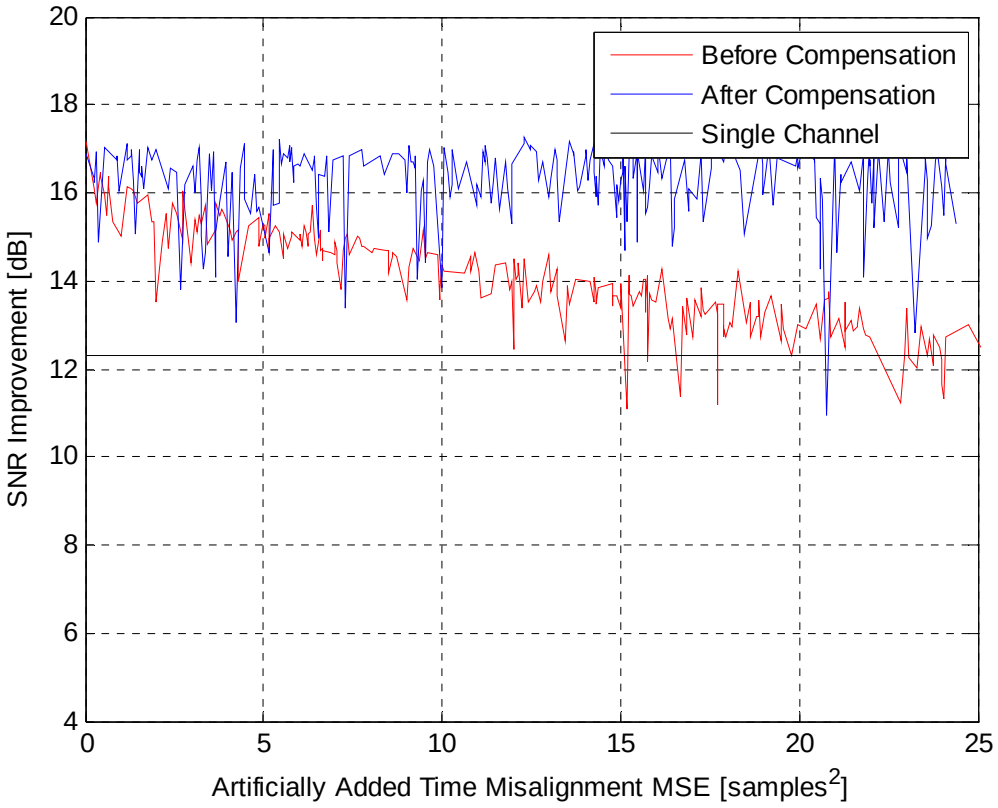

Figure H-21 SNR Improvement for 32 Microphones, -10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

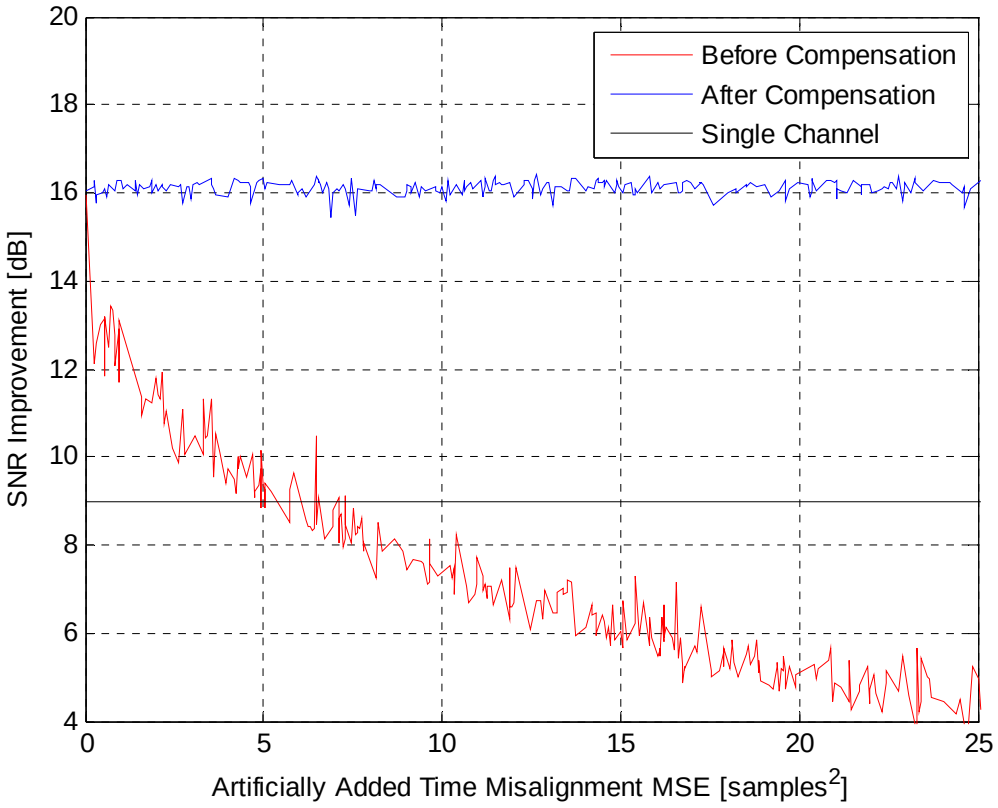

Figure H-22 SNR Improvement for 32 Microphones, 0 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

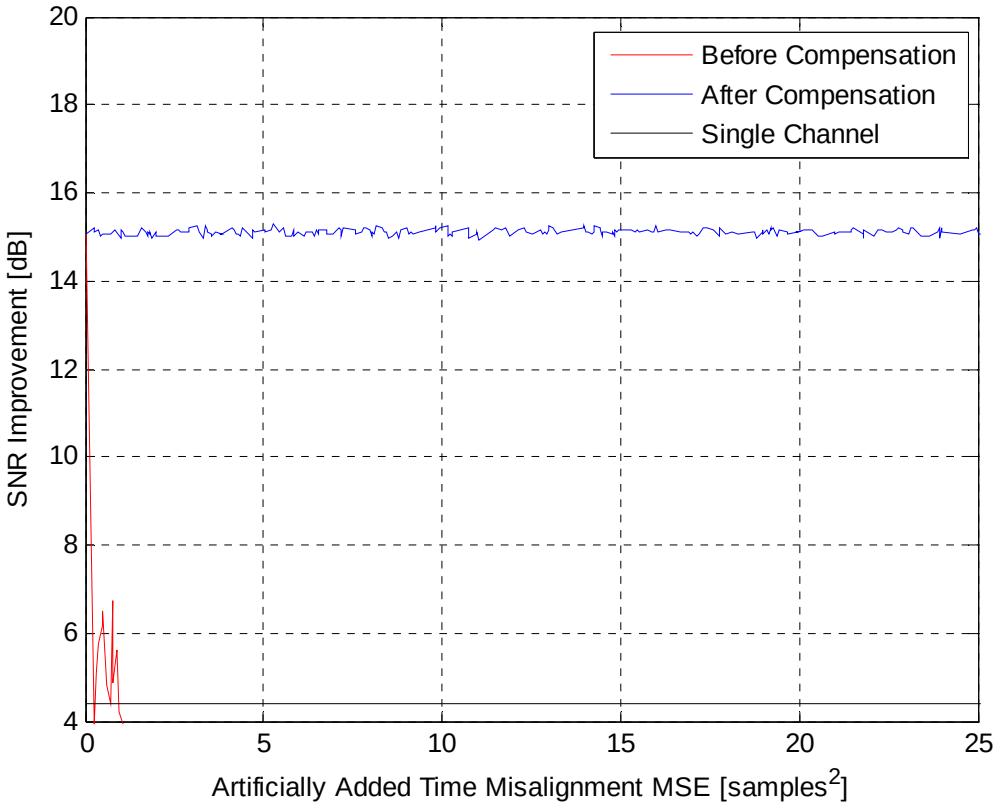

Figure H-23 SNR Improvement for 32 Microphones, 10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

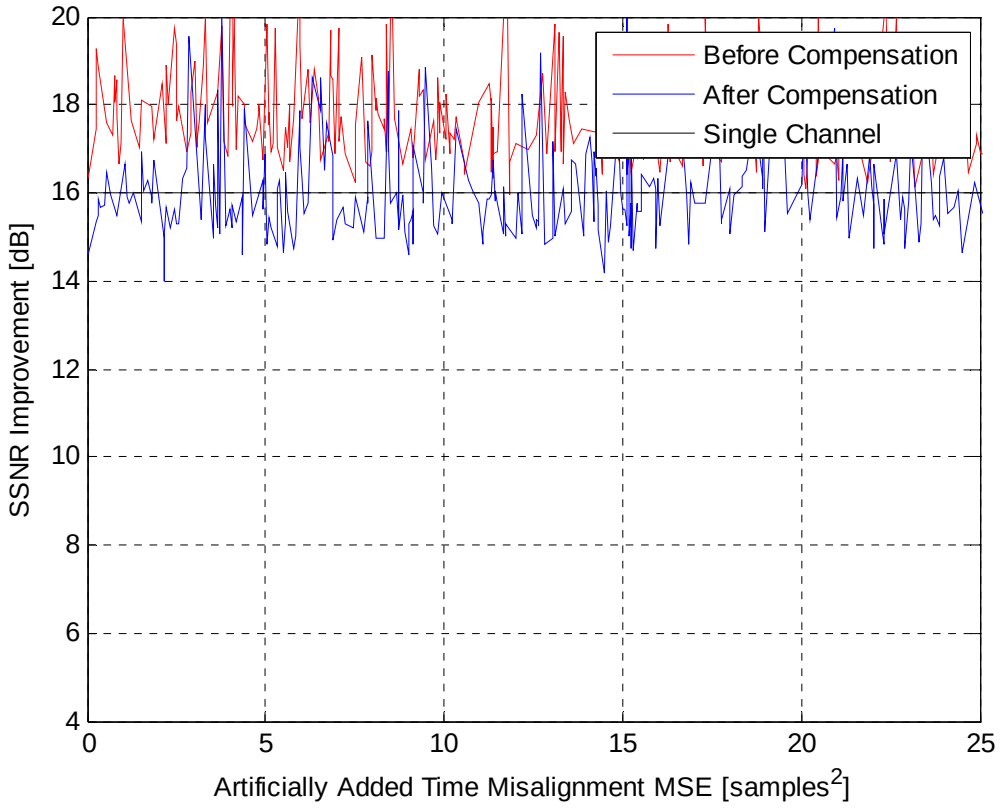

Figure H-24 SSNR Improvement for 32 Microphones, -20 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

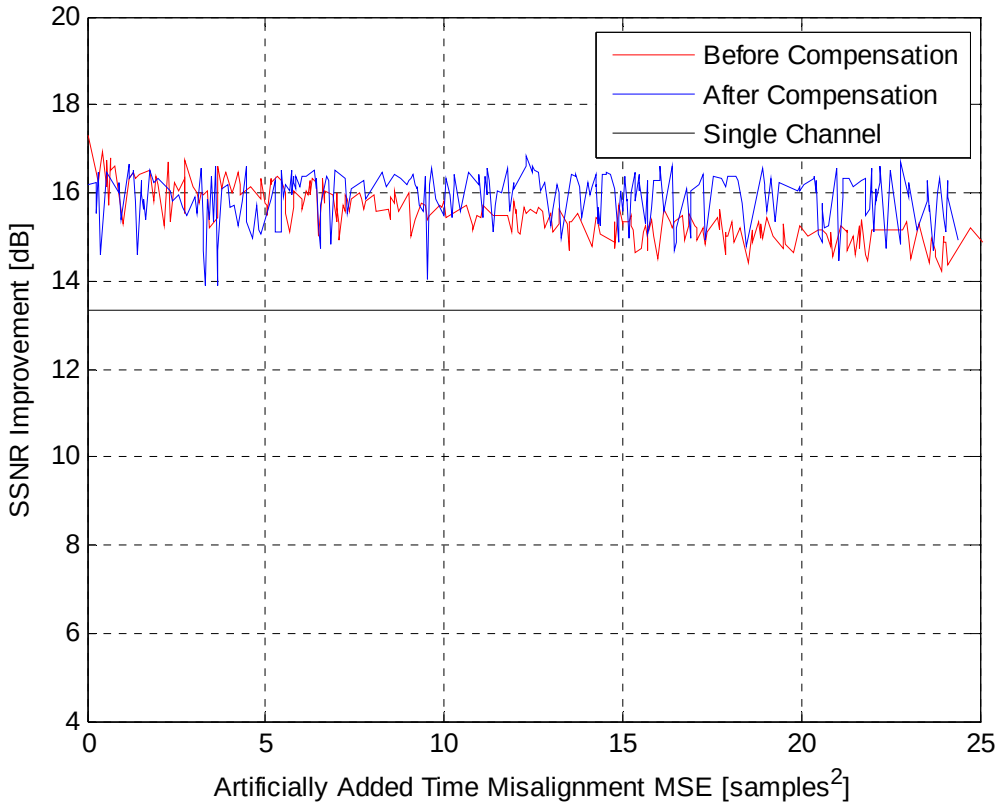

Figure H-25 SSNR Improvement for 32 Microphones, -10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

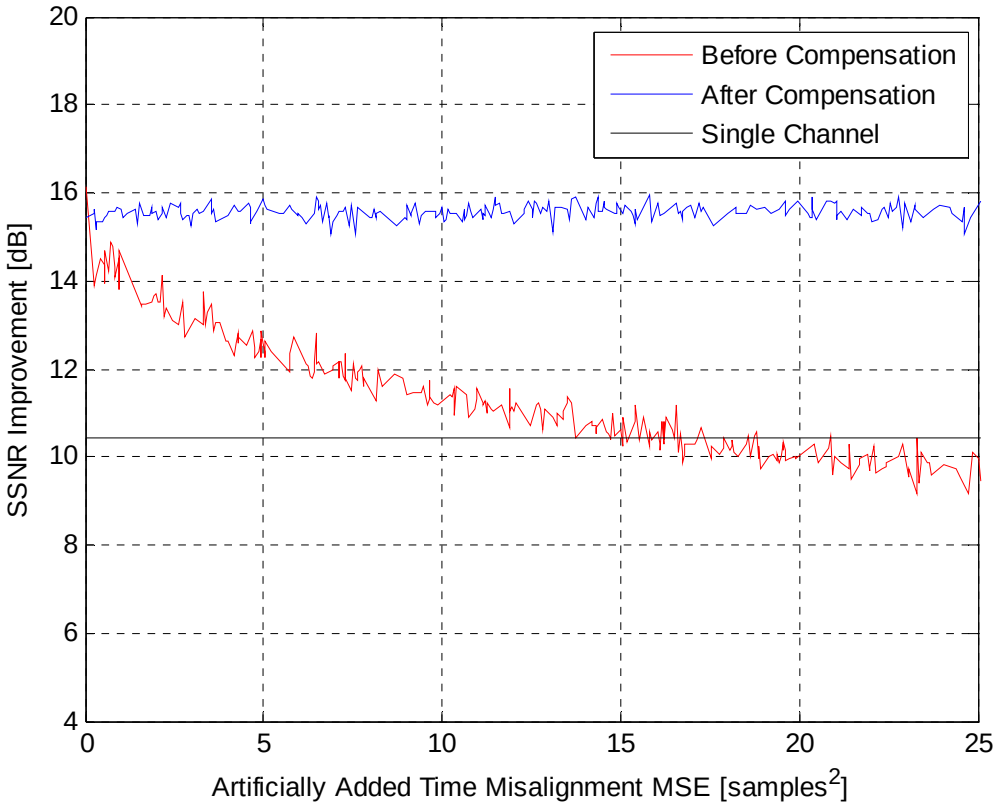

Figure H-26 SSNR Improvement for 32 Microphones, 0 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

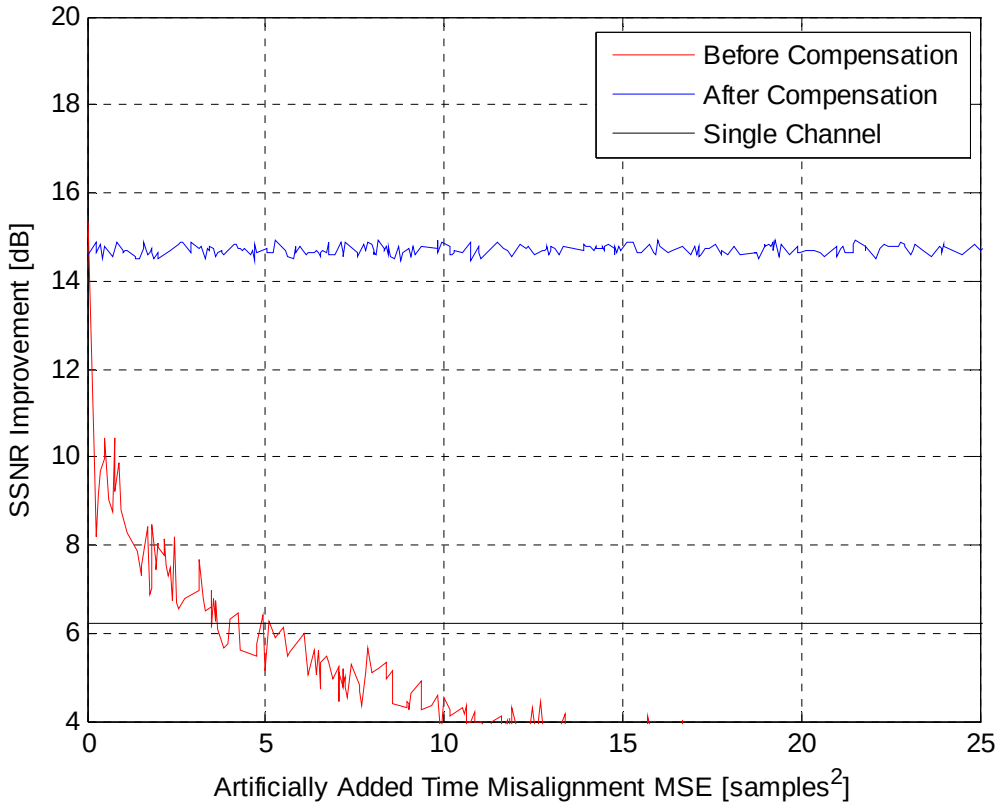

Figure H-27 SSNR Improvement for 32 Microphones, 10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation Before and After Artificial Time Misalignment Compensation

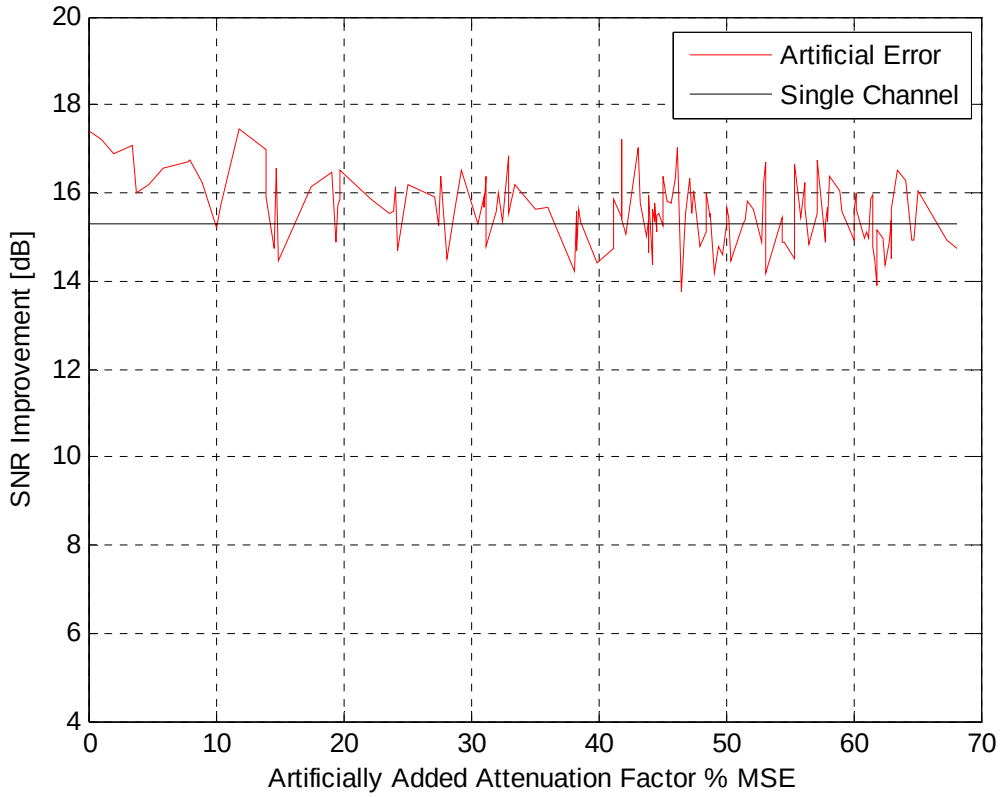

Figure H-28 SNR Improvement for 32 Microphones, -20 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

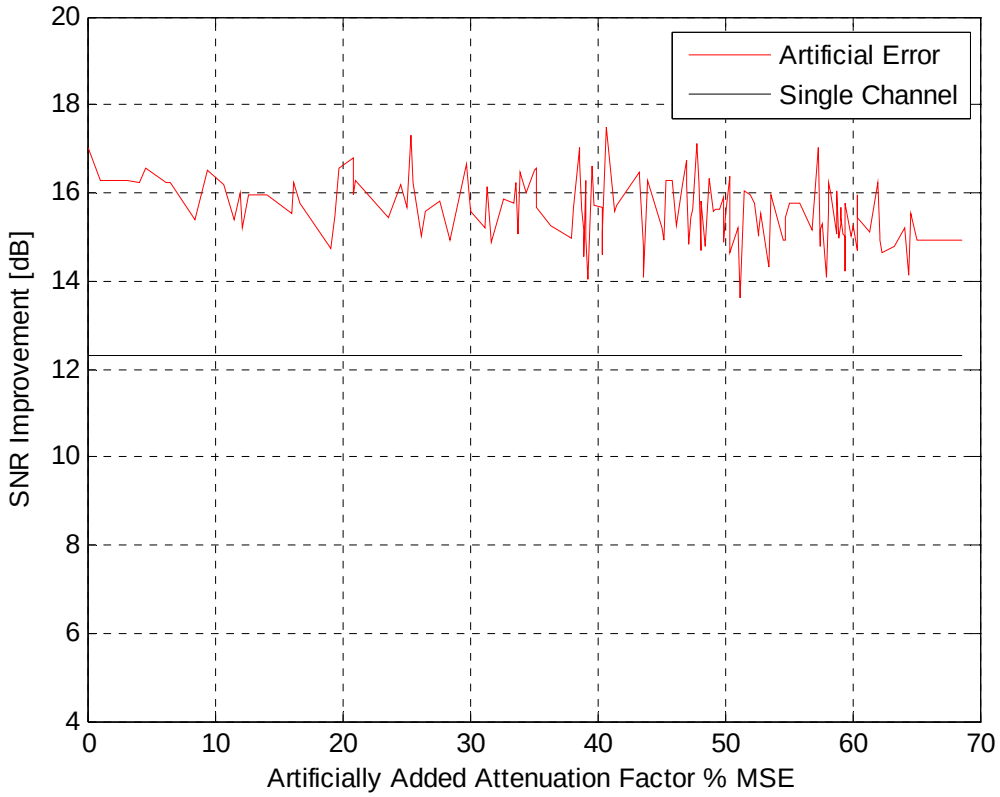

Figure H-29 SNR Improvement for 32 Microphones, -10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

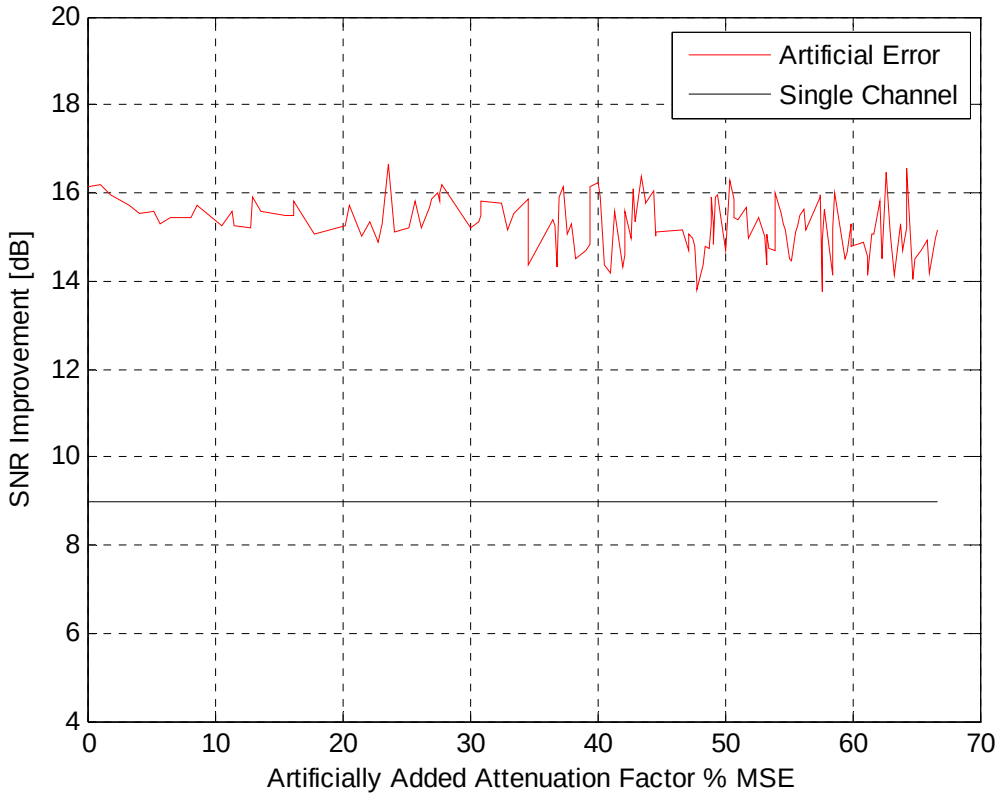

Figure H-30 SNR Improvement for 32 Microphones, 0 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

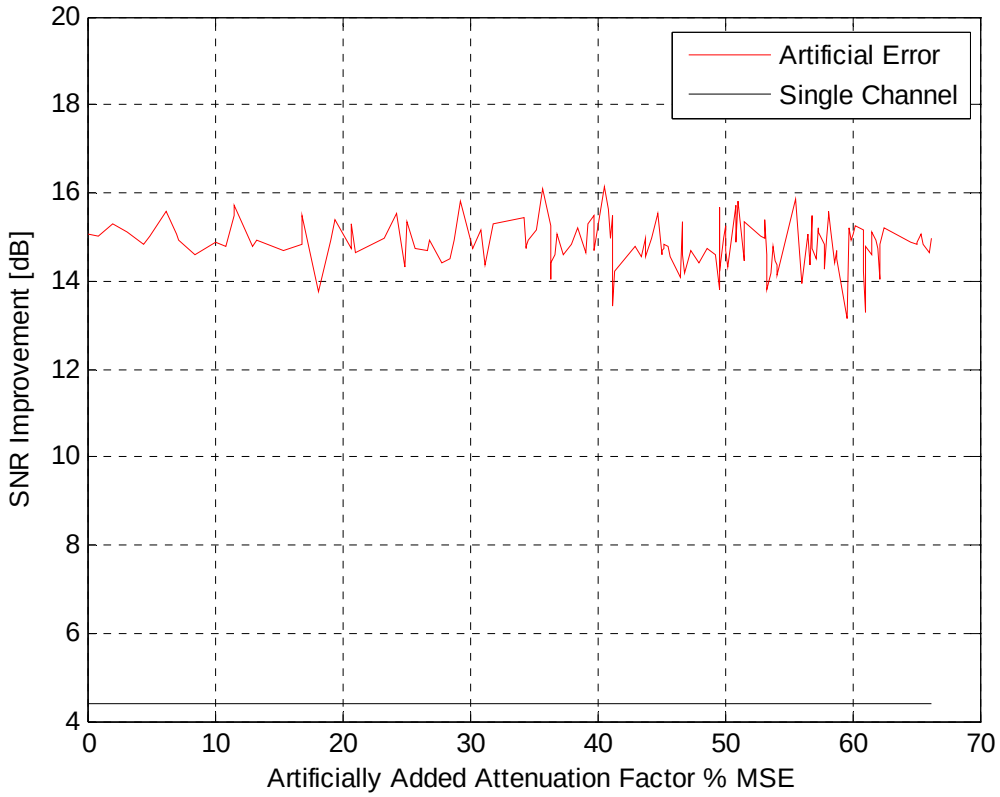

Figure H-31 SNR Improvement for 32 Microphones, 10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

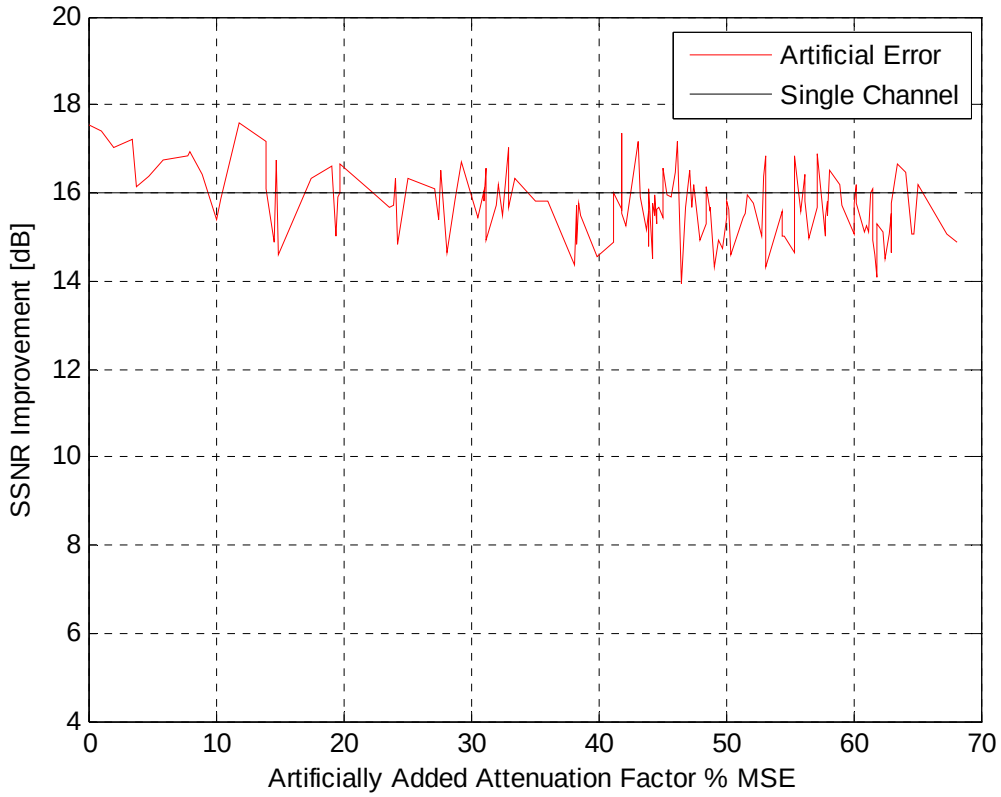

Figure H-32 SSNR Improvement for 32 Microphones, -20 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

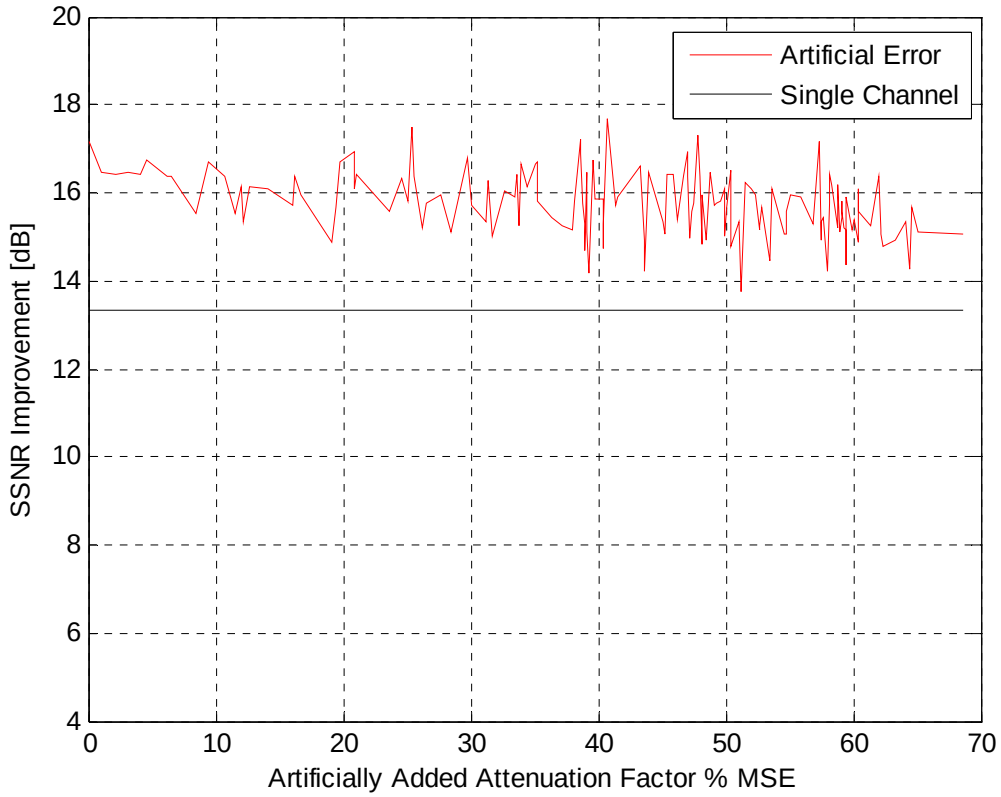

Figure H-33 SSNR Improvement for 32 Microphones, -10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

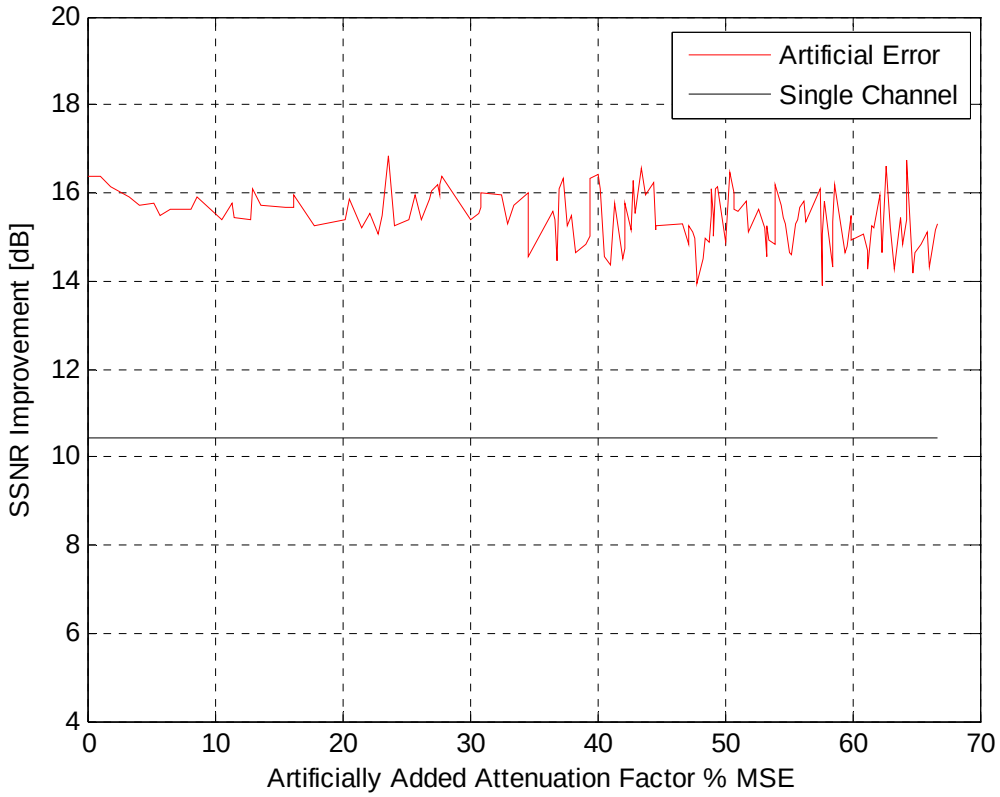

Figure H-34 SSNR Improvement for 32 Microphones, 0 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors

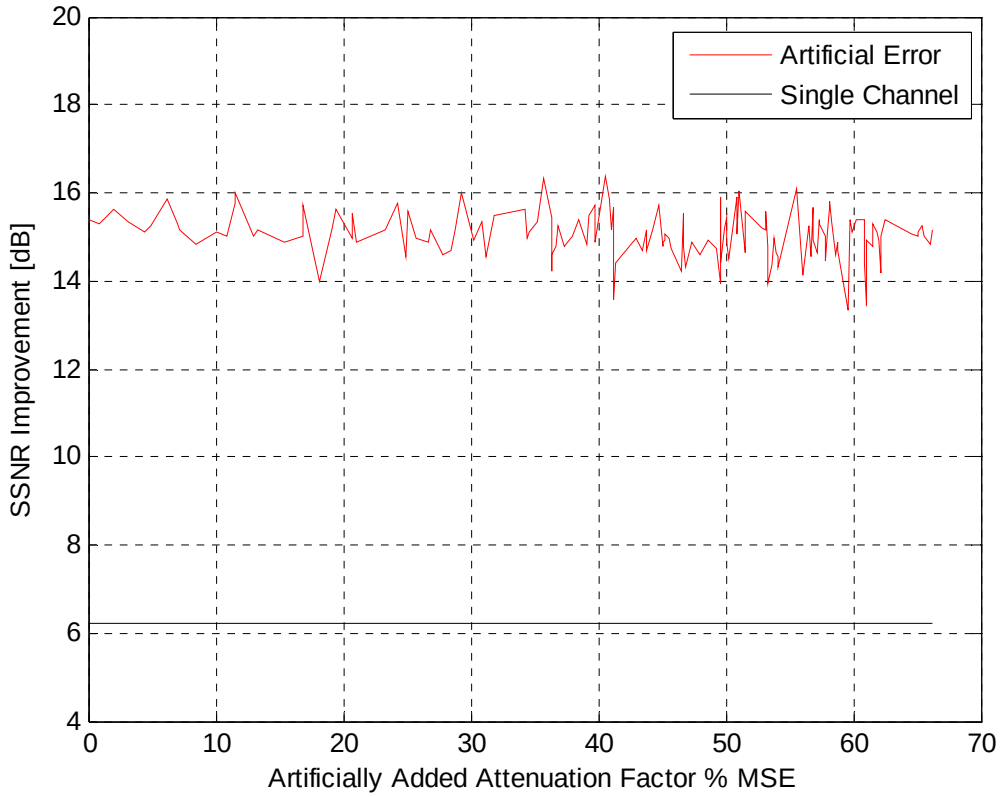

Figure H-35 SSNR Improvement for 32 Microphones, 10 dB Input SNR, and Unity Attenuation Factors using Log-Spectral Amplitude Estimation (LSA) with Spectral Phase Estimation due to Artificial Error in Attenuation Factors