

Marquette University

e-Publications@Marquette

Electrical and Computer Engineering Faculty
Research and Publications

Electrical and Computer Engineering,
Department of

2-2012

Distributed Multichannel Speech Enhancement with Minimum Mean-square Error Short-time Spectral Amplitude, Log-spectral Amplitude, and Spectral Phase Estimation

Marek B. Trawicki

Marquette University, marek.trawicki@marquette.edu

Michael T. Johnson

Marquette University, michael.johnson@marquette.edu

Follow this and additional works at: https://epublications.marquette.edu/electric_fac



Part of the [Computer Engineering Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Trawicki, Marek B. and Johnson, Michael T., "Distributed Multichannel Speech Enhancement with Minimum Mean-square Error Short-time Spectral Amplitude, Log-spectral Amplitude, and Spectral Phase Estimation" (2012). *Electrical and Computer Engineering Faculty Research and Publications*. 56.
https://epublications.marquette.edu/electric_fac/56

Marquette University

e-Publications@Marquette

Electrical and Computer Engineering Faculty Research and Publications/College of Engineering

This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.

Signal Processing, Vol. 92, No. 2 (February 2012): 345-356. [DOI](#). This article is © Elsevier and permission has been granted for this version to appear in [e-Publications@Marquette](#). Elsevier does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Elsevier.

Distributed Multichannel Speech Enhancement with Minimum Mean-Square Error Short-Time Spectral Amplitude, Log-Spectral Amplitude, And Spectral Phase Estimation

Marek B. Trawicki

Marquette University, Department of Electrical and Computer Engineering, Speech and Signal Processing Laboratory, Milwaukee, WI

Michael T. Johnson

Marquette University, Department of Electrical and Computer Engineering, Speech and Signal Processing Laboratory, Milwaukee, WI

Abstract

In this paper, the authors present optimal multichannel frequency domain estimators for minimum mean-square error (MMSE) short-time spectral amplitude (STSA), log-spectral amplitude (LSA), and spectral phase estimation in a widely distributed microphone configuration. The estimators utilize Rayleigh and Gaussian statistical models for the speech prior and noise likelihood with a diffuse noise field for the surrounding environment. Based on the Signal-to-Noise Ratio (SNR) and Segmental Signal-to-Noise Ratio (SSNR) along with the Log-Likelihood Ratio (LLR) and Perceptual Evaluation of Speech Quality (PESQ) as objective metrics, the multichannel LSA estimator decreases background noise and speech distortion and increases speech quality compared to the baseline single channel STSA and LSA estimators, where the optimal multichannel spectral phase estimator serves as a significant quantity to the improvements, and demonstrates robustness due to time alignment and attenuation factor estimation. Overall, the optimal distributed microphone spectral estimators show strong results in noisy environments with application to many consumer, industrial, and military products.

Keywords

Acoustic arrays, Speech enhancement, Amplitude estimation, Phase estimation, Parameter estimation

1. Introduction

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 noisy conditions. In order to decrease background noise and speech distortion and increase speech quality, which are measured by Signal-to-Noise Ratio (SNR) and Segmental Signal-to-Noise Ratio (SSNR) [1] along with the Log-Likelihood Ratio (LLR) [2] and Perceptual Evaluation of Speech Quality (PESQ) [3] as objective metrics [4], researchers have utilized multichannel (dual, array, and distributed) microphones to exploit all available acoustic and spatial information of the speech and noise sources [5]. While single channel microphone configurations require the speakers to be relatively close to the microphone and dual channel microphone configurations involve a reference noise microphone [6], microphone array [7] configurations necessitate close-spacing of the microphones and *a priori* knowledge of the array geometry with the distances between individual array elements being small enough to allow for spatial signal processing techniques (e.g., beamforming) without aliasing and justify assumptions of noise correlation across the channels [6], [8], [9], [10], [11], [12]. By comparison, there has been relatively little research for distributed microphone configurations, where the microphones are spread throughout a large area of interest with unknown spacing and geometry and array assumptions do not hold anymore. 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.

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., offices, cafeterias, and 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 [8]. The magnitude-squared coherence (MSC) function, $C_{ij}(f) = \text{sinc}^2(2\pi f d_{ij}/c)$ [7], where d_{ij} is the distance between channels i and j , and c is the speed of sound, can be used to give an indication of noise correlation for a given microphone spacing. Since the primary energy of speech is mainly concentrated in the 300–3000 Hz frequency range, the MSC function suggests as examples that an assumption of incoherent noise ($C < 0.1$) is justified for microphone spacing above ~ 14 cm and an assumption of coherent noise ($C > 0.9$) is justified only for microphone spacing below

~ 0.4 cm, which is less than the distances in a typical array. For distributed microphone spectral amplitude and spectral phase estimators derived in this work, the noise field is assumed to be a diffuse noise field, which allows for estimation of the noise statistics at each of the corresponding microphones since the coherence is approximately small for high frequencies outside the primary energy of speech.

In this paper, the goal is to derive optimal multichannel frequency domain estimators of the short-time spectral amplitude (STSA) and log spectral amplitude (LSA) with spectral phase for distributed microphone configurations using a minimum mean-square error (MMSE) estimation criterion. This theoretical work can be viewed as an extension of the seminal work for single channel STSA [9], LSA [10], and spectral phase [9], [10] estimation by Ephraim and Malah. Fundamentally, the statistical estimation [13], [14] approach employs the same Rayleigh and Gaussian models for the speech prior and noise likelihood and diffuse noise field as Lotter et al. [11] but generalizes the derivations to estimate the spectral amplitude of the source signal, not the source signal as seen at each individual microphone. Besides the multichannel STSA and LSA estimators for distributed microphone configurations, the key contribution is the development of the multichannel spectral phase estimator since a substantially improved estimate of the true source spectral phase can be calculated from multiple channels.

The remainder of this paper is organized into the following sections: system and models (Section 2), spectral amplitude estimation (Section 3), log-spectral amplitude estimation (Section 4), spectral phase estimation (Section 5), experiments and implementation (Section 6), experimental results (Section 7), and conclusion (Section 8).

2. System and models

In a distributed microphone configuration, multiple microphones $i \in [1, \dots, M]$ capture the attenuated and time-delayed coherent clean source signals $c_i s(t - \tau_i)$ corrupted by additive and uncorrelated noises $n_i(t)$ to produce the noisy signals $y_i(t)$. Assuming the system can accurately time align the M noisy observations (investigated in Section 7), the time domain multichannel microphone model is

$$(1) y_i(t) = c_i s(t) + n_i(t),$$

where $s(t)$ is the true source signal and $c_i \in [0, 1]$ are time-invariant attenuation factors. This model corresponds to an assumption of a spatially stationary source signal. In the frequency-domain, (1) is expressed in a frame-by-frame basis as

$$(2) Y_i(l, k) = c_i S(l, k) + N_i(l, k) R_i(l, k) e^{j\vartheta_i(l, k)} = c_i A(l, k) e^{j\alpha(l, k)} + N_i(l, k),$$

where k represents the frequency bin for each microphone i and frame l . To simplify the notation, (2) is rewritten without the explicit dependencies as

$$(3) R_i e^{j\vartheta_i} = c_i A e^{j\alpha} + N_i,$$

where R_i and A represent the noisy and clean spectral amplitudes, respectively, ϑ_i and α represent the noisy and clean spectral phases, respectively, and N_i represents the spectral noise. The fundamental goal is to determine the best estimate of the spectral amplitude A and spectral phase α .

Rayleigh distributions are assumed for the speech prior likelihood:

$$(4) p(A, \alpha) = \frac{A}{\pi \sigma_s^2} \exp\left(-\frac{A^2}{\sigma_s^2}\right),$$

and Gaussian distributions are assumed for the noise likelihood:

$$(5) p(Y_i|A, \alpha) = \frac{1}{\pi \sigma_{N_i}^2} \exp\left(-\frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right),$$

where σ_S^2 and $\sigma_{N_i}^2$ are the speech and noise spectral variances, respectively, and $A > 0$. 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 Y_1, \dots, Y_M written as a product of the independent noisy spectral observations given by

$$(6) p(Y_1, \dots, Y_M|A, \alpha) = \prod_{i=1}^M p(Y_i|A, \alpha) = \prod_{i=1}^M \frac{1}{\pi \sigma_{N_i}^2} \exp\left(-\sum_{i=1}^M \frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right).$$

3. Spectral amplitude estimation

Under the given statistical models and following the same method as in [\[11\]](#), the MMSE estimate of the STSA is

$$(7) \hat{A}_{STSA} = E[A|Y_1, \dots, Y_M] = \frac{\int_0^\infty \int_0^{2\pi} A p(Y_1, \dots, Y_M|A, \alpha) p(A, \alpha) d\alpha dA}{\int_0^\infty \int_0^{2\pi} p(Y_1, \dots, Y_M|A, \alpha) p(A, \alpha) d\alpha dA}.$$

Through substitution of the statistical models of [\(4\)](#), [\(6\)](#) into [\(7\)](#), the closed-form solution (see [Appendix A](#) for details) for \hat{A}_{STSA} is

$$(8) \hat{A}_{STSA} = \Gamma(1.5) \left(\frac{\sigma_S^2}{1 + \sum_{i=1}^M \xi_i}\right)^{1/2} \exp\left(-\frac{v}{2}\right) \left[(1 + v)I_0\left(\frac{v}{2}\right) + vI_1\left(\frac{v}{2}\right)\right],$$

with

$$(9) v = \frac{\left|\sum_{i=1}^M (\sqrt{\xi_i}/\sigma_{N_i}) Y_i\right|^2}{1 + \sum_{i=1}^M \xi_i},$$

where $I_0(\cdot)$ and $I_1(\cdot)$ denote the modified Bessel functions of the first kind of the 0th order and 1st order, respectively, and ξ_i is the *a priori* SNR. In more qualitative terms, v is an SNR weighted sum of the noisy spectral observations Y_i that is normalized by the sum of the *a priori* SNR ξ_i . For the case of $M = 1$, it can be seen that the multichannel STSA estimator in [\(8\)](#) simplifies to the single channel STSA estimator in [\[9\]](#). With rescaling of the attenuation factors to make $c_i = 1$ at a specific reference channel i , [\(8\)](#) reduces to the multichannel STSA estimator in [\[11\]](#) for estimating the spectral amplitude A_i at each microphone i . The only difference between the multichannel STSA estimator given in [\[11\]](#) by Lotter et al. and the multichannel STSA estimator given in [\(8\)](#) is that [\(8\)](#) is an estimate of the original source STSA, not an estimate of the original source STSA at a particular channel.

4. Log-spectral amplitude estimation

To obtain a more perceptually relevant criterion [\[10\]](#), the MMSE estimate from [Section 3](#) is extended as

$$(10) \hat{A}_{LSA} = \exp(E[\ln(A)|Y_1, \dots, Y_M]) = \exp(E[Z|Y_1, \dots, Y_M]),$$

where

$$(11) E[Z|Y_1, \dots, Y_M] = \frac{d}{d\mu} [\Phi_{Z|Y_1, \dots, Y_M}(\mu)]|_{\mu=0},$$

is the moment generating function:

$$(12) \Phi_{Z|Y_1, \dots, Y_M}(\mu) = \frac{\int_0^\infty \int_0^{2\pi} A^\mu p(Y_1, \dots, Y_M|A, \alpha) p(A, \alpha) d\alpha dA}{\int_0^\infty \int_0^{2\pi} p(Y_1, \dots, Y_M|A, \alpha) p(A, \alpha) d\alpha dA}.$$

By substitution of the statistical models of [\(4\)](#), [\(6\)](#) into [\(12\)](#), the closed-form solution (see [Appendix B](#) for details) for $\Phi_{Z|Y_1, \dots, Y_M}(\mu)$ is

$$(13) \Phi_{Z|Y_1, \dots, Y_M}(\mu) = \frac{\Gamma((\mu/2)+1)}{(1/\lambda)^{\mu/2}} {}_1F_1\left(-\frac{\mu}{2}; 1; -v\right),$$

where

$$(14) \frac{1}{\lambda} = \frac{1}{\sigma_S^2} + \sum_{i=1}^M \frac{c_i^2}{\sigma_{N_i}^2},$$

and ${}_1F_1(\cdot; \cdot; \cdot)$ denotes the confluent hypergeometric function as described by equation 9.210 in [\[15\]](#).

To complete the derivation of the multichannel LSA estimator, it is necessary to perform differentiation and then exponentiation on [\(13\)](#). After exponentiation of the derivative terms, the closed-form solution (see [Appendix B](#) for details) for \hat{A}_{LSA} is written as

$$(15) \hat{A}_{LSA} = \left(\frac{\sum_{i=1}^M \xi_i / \gamma_i}{\sum_{i=1}^M c_i^2 / R_i^2} \right)^{1/2} \left(\frac{\left| \sum_{i=1}^M (\sqrt{\xi_i} / \sigma_{N_i})^{Y_i} \right|}{1 + \sum_{i=1}^M \xi_i} \right) \exp \left(\frac{1}{2} \int_v^\infty \frac{e^{-t}}{t} dt \right),$$

where v is defined in [\(9\)](#). For the case of $M = 1$, it can be seen that the multichannel LSA estimator in [\(15\)](#) simplifies to the single channel LSA estimator in [\[10\]](#). As with the estimate of the multichannel STSA estimator [\(8\)](#), the noisy spectral observations Y_i in the multichannel LSA estimator [\(15\)](#) that contain more clean spectral observations S than noise spectral observations N_i will be weighted higher than noisy observations Y_i that contain less clean spectral observations S than noise spectral observations N_i .

5. Spectral phase estimation

The estimation of spectral phase is also an important component for constructing the enhanced true source signal. As shown for the single channel MMSE spectral phase estimator in [\[9\]](#), the MMSE estimation of the complex exponential estimator $e^{j\alpha}$ results in a non-unity modulus, which produces an altered and a non-optimal estimate of the STSA and LSA. In order to prevent the optimal spectral phase estimator from affecting the optimal STSA and LSA estimates, the approach taken in this work is the same constrained optimization formulation as in [\[9\]](#) according to the minimum:

$$(16) \min_{e^{j\hat{\alpha}}} E[|e^{j\alpha} - e^{j\hat{\alpha}}|^2]$$

$$\text{subject to } |e^{j\hat{\alpha}}| = 1,$$

where the amplitude of the complex exponential is constrained to have unity modulus. Through the Lagrange multiplier optimization method, (16) is reformulated as

$$(17) \min_{g, \rho} E[|e^{j\alpha} - g|^2 | Y_1, \dots, Y_M] + \rho(|g| - 1)$$

$$\text{subject to } |g| = 1,$$

with

$$(18) g = e^{j\alpha} = g_R + jg_I,$$

and ρ serving as the Lagrange multiplier.

Under the formulation in (17), the constrained MMSE spectral phase solution is

$$(19) \hat{\alpha} = \tan^{-1}(g_I/g_R).$$

From (19), the key relationship between the real and imaginary components in (18) is

$$(20) \frac{g_I}{g_R} = \frac{E[\sin\alpha | Y_1, \dots, Y_M]}{E[\cos\alpha | Y_1, \dots, Y_M]}.$$

By solving and simplifying (20) with the attenuated spectral amplitude $A_i = c_i A$ and attenuated spectral variance $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$, the closed-form solution for $\hat{\alpha}$ in (19) (see Appendix C for details) is

$$(21) \hat{\alpha} = \tan^{-1} \left(\frac{\sum_{i=1}^M (\sqrt{\xi_i}/\sigma_{N_i}) \text{Im}(Y_i)}{\sum_{i=1}^M (\sqrt{\xi_i}/\sigma_{N_i}) \text{Re}(Y_i)} \right),$$

which is a quotient of two weighted sums of the noisy spectral observations [16]. As with the multichannel STSA and LSA estimators given in (8), (15), the multichannel spectral phase estimator simplifies in (21) to the well-known single channel spectral phase MMSE estimator in [9] for the case of $M = 1$, which is simply the noisy spectral phase ϑ .

6. Experiments and implementation

6.1. Enhancement

To evaluate the proposed optimal multichannel STSA, LSA, and spectral phase estimators derived in (8), (15), and (21), distributed multiple microphone noisy signals were simulated using the TIMIT [17] and NOISEX [18] corpora. The 10 speech signals, which averaged 2.4 ± 0.5 s in length, were sampled at 16 kHz and corrupted by white, pink, and babble noises with input SNRs of -10 dB, 0 dB, and +10 dB, respectively, for 1–24 microphones and two different attenuation factor configurations, namely constant ($c_i = 1$) and linearly decreasing ($c_i = (M - i + 1)/M$), where M and i represent the total number of microphones and a specific microphone index, respectively. The noisy speech signals were truncated to produce an equal number of

samples in each frame. For each of the noisy spectral observations Y_i , the analysis conditions consisted of frames of 256 samples (16 ms) with 50% overlap between the corresponding frames using Hanning windows. Noise estimation was performed on 5 initial silence frames without any subsequent updating of the time series or spectrum. The decision-directed (DD) [9] smoothing approach was utilized to recursively estimate the *a priori* SNR ξ_i as

$$(22) \hat{\xi}_i = \frac{\sigma_{S_i}^2}{\sigma_{N_i}^2} = \frac{c_i^2 \sigma_S^2}{\sigma_{N_i}^2} = \alpha_{SNR} \hat{c}_i^2 \frac{\hat{A}^{2(l-1)}}{\sigma_{N_i}^2} + (1 - \alpha_{SNR}) P[\gamma_i(l) - 1],$$

and the *a posteriori* SNR γ_i was calculated as

$$(23) \gamma_i = R_i^2 / \sigma_{N_i}^2,$$

for each channel i and frame l with $\alpha_{SNR} = 0.98$ using thresholds of $\xi_{min} = 10^{-25/10}$ for perceptual reasons [19] and $\gamma_{max} = 40$ (implemented as a floor on $\sigma_{N_i}^2$) to avoid numeric overflows.

6.2. Spectral phase estimation

In order to evaluate the derived multichannel spectral phase estimator $\hat{\alpha}$ given in (21), which is a central contribution to the enhancement of the noisy spectral observations Y_i , experiments were run comparing the SSNR of the new multichannel spectral phase estimator to the SSNR of the standard single channel spectral phase estimator, which is simply the noisy spectral phase ϑ of the reference channel $i = 1$. The spectral phase estimation experiments were implemented for a 24 microphone scenario in white noise at 0 dB input SNR with constant attenuation factors.

6.3. Time alignment

For multiple microphones, initial time alignment is required of the channels. To implement alignment, time delays can be estimated through a variety of methods, which are similar to Time-Difference-of-Arrival (TDOA) methods used for source localization [20]. 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.

To evaluate the impact of artificially added misalignment as well as the effectiveness of the selected time alignment method, the noisy observations were time shifted by a random number of samples selected from a zero-mean Gaussian distribution with gradually increasing variance. The time alignment experiments were implemented for a 24 microphone scenario in white noise at 0 dB input SNR with constant attenuation factors.

6.4. 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 in (22). 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 i . 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 [21] 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.

For the experiments, the reference microphone is defined as $i = 1$ with $c_1 = 1$. The remaining attenuation factors c_i are estimated from the signal variances of the noisy observations y_i under the assumed independence of the speech s and noises n_i using

$$(24) \hat{c}_i = \sqrt{\sigma_{y_i}^2 - \sigma_{n_i}^2} / \sigma_s = \sqrt{\sigma_{y_i}^2 - \sigma_{n_i}^2} / \sqrt{\sigma_{y_1}^2 - \sigma_{n_1}^2},$$

which is a ratio of two estimated clean signal powers. 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 enhancement was evaluated by adding random error selected from a zero-mean Gaussian distribution with increasing variance. 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 24 microphone scenario in white noise at 0 dB input SNR with constant attenuation factors.

7. Experimental results

The baseline methods used for comparison of the results are the multichannel STSA estimator with single channel (noisy) spectral phase estimator (i.e., equivalent to the method of Lotter et al. [11]) as well as the single channel STSA estimator with single channel (noisy) spectral phase estimator [9] and single channel LSA estimator with single channel (noisy) spectral phase estimator [10], which are equivalent to the newly proposed optimal estimators for $M = 1$. Objective measures of SSNR, LLR, and PESQ were utilized to measure the noise reduction, speech distortion, and overall quality of the enhanced signals, respectively, which were reconstructed using the overlap-add technique, at the arbitrary reference microphone $i = 1$ with $c_1 = 1$. At input SNRs of -10 dB, 0 dB, and $+10$ dB, the input LLR and input PESQ were 1.70, 1.64, and 1.34 and 1.15, 1.63, and 2.28, respectively. Due to the frequency analysis performed on each of the microphones along with estimation of the noise, *a priori* and *a posteriori* SNRs, and attenuation factors independently across the microphones, the computational complexity for the multichannel channel estimators is simply M times the corresponding single channel estimator.

7.1. Enhancement

[Fig. 1](#), [Fig. 2](#), [Fig. 3](#) show the SSNR improvement, LLR output, and PESQ output, respectively, as a function of the number of microphone channels in the array across the noises (white, pink, and babble) at -10 dB, 0 dB, and $+10$ dB input SNRs with constant and linear attenuation factors, where LLR has range of 0 – 2 (lower scores indicate better performance) and PESQ has range of 0.5 – 4.5 (higher scores indicate better performance).

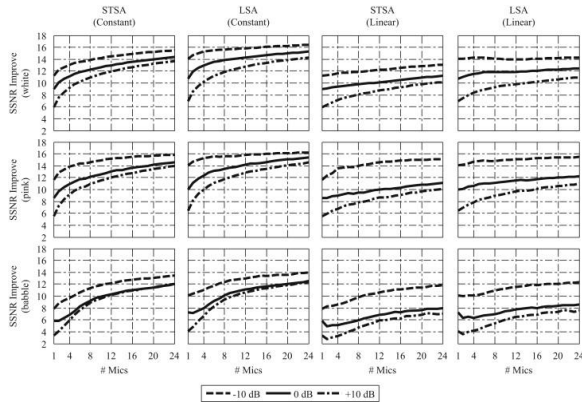


Fig. 1. SSNR Improvement for multichannel STSA and LSA estimators with multichannel spectral phase estimator for constant and linear attenuation factors.

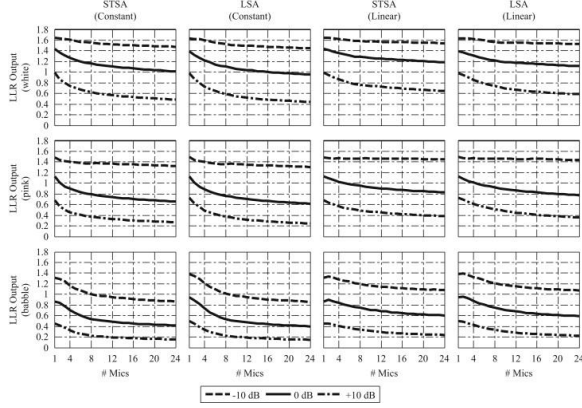


Fig. 2. LLR output for multichannel STSA and LSA estimators with multichannel spectral phase estimator for constant and linear attenuation factors.

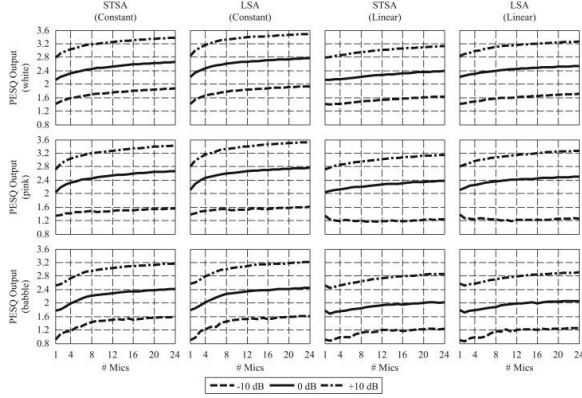


Fig. 3. PESQ output for multichannel STSA and LSA estimators with multichannel spectral phase estimator for constant and linear attenuation factors.

From the results, the constant attenuation factor scenario offered the most substantial increases in noise reduction, decreases in speech distortion, and increases in overall speech quality across the noises and input SNR levels for varying number of microphones. In terms of SSNR improvements, the multichannel LSA estimator with multichannel spectral phase estimator provided 2–4 dB (–10 dB), 5–6 dB (0 dB), and 7–8 dB (+10 dB) increase in noise reduction over the corresponding single channel baseline LSA estimator. Typically, the multichannel LSA estimator with multichannel spectral phase estimator had a 1–2 dB increase over the multichannel STSA estimator with multichannel spectral phase estimator with largest gains for white noise. For the LLR outputs, there were decreases in speech distortion of 0.2–0.6 (–10 dB), 0.4–0.6 (0 dB), and 0.3–0.6 (+10 dB) for the multichannel LSA estimator with multichannel spectral phase estimator over the single channel

baseline LSA estimator. In general, the multichannel STSA and LSA estimators with multichannel spectral phase estimator had similar decreasing curves with the best LLR outputs for babble noise. With PESQ outputs, the speech quality improved by 0.5–0.6 (–10 dB), 0.6–0.8 (0 dB), and 0.3–0.6 (+10 dB) multichannel LSA estimator with multichannel spectral phase estimator. The multichannel STSA estimator with multichannel spectral phase estimator had nearly identical PESQ outputs as the multichannel LSA estimator with multichannel spectral phase estimator showing the most pronounced gains involving white and pink noises. By comparing the multichannel STSA and LSA estimators with multichannel spectral phase estimator for linear attenuation factors, the SSNR improvement, LLR output, and PESQ output results were slightly worse than the analogous results for constant attenuation factors, which is a reasonable outcome since the noisy microphone channels with linear attenuation factors are not the same distant away from the source signal and consequently do not provide equally useful information as with the constant attenuation factors.

7.2. Spectral phase estimation

[Fig. 4](#) presents the SSNR improvement difference between the single channel and multichannel spectral phase estimators with the multichannel LSA estimator as a function of the number of microphone channels in the array across the noises (white, pink, and babble) at –10 dB, 0 dB, and +10 dB input SNRs with constant and linear attenuation factors.

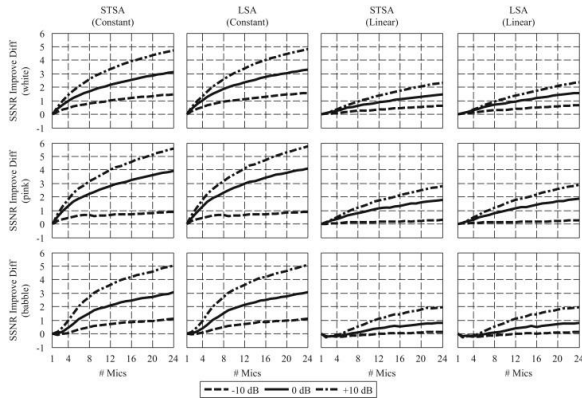


Fig. 4. SSNR improvement difference between single channel spectral phase and multichannel spectral phase estimators with multichannel STSA estimator and multichannel LSA estimator for constant and linear attenuation factors.

Based on the results, the derived multichannel spectral estimator in [\(21\)](#) surpassed the baseline single channel (noisy) spectral phase estimator by upwards of 5–6 dB (–10 dB), 3–4 dB (0 dB), and 1–2 dB (+10 dB) across the noises for the multichannel STSA and LSA estimators with constant attenuation factors. By contrast, there is less SSNR improvement between the multichannel and single channel spectral phase estimators using linear attenuation factors. Specifically, the SSNR improvement difference ranged from 2–3 dB (–10 dB), 1–2 dB (0 dB), and 0–1 dB (+10 dB) across the noises for the multichannel STSA and LSA estimators. In general, there is consistent SSNR improvement due to the multichannel spectral over the baseline single channel (noisy) spectral phase estimator with increasing number of microphone channels, which constitutes a significant portion of the overall improvement obtained when using all of the available acoustic and spatial information from the noisy signals in the surrounding environment.

7.3. Time alignment

[Fig. 5](#) illustrates the impact of artificial misalignment, as measured by Mean-Square Error (MSE), and corresponding automatic time alignment of a 24 microphone channel configuration for the multichannel LSA

estimator with multichannel spectral phase estimator across white noise at 0 dB input SNR with constant attenuation factors.

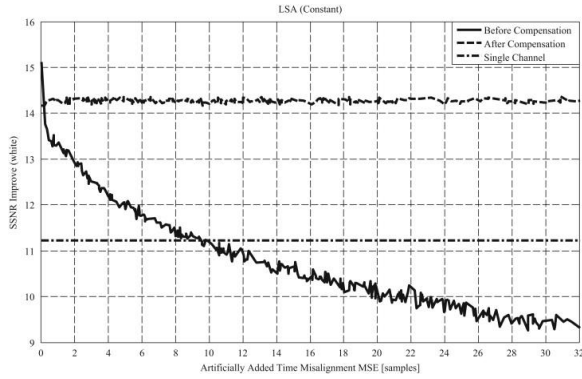


Fig. 5. SSNR improvement for multichannel LSA estimator with multichannel spectral phase estimator before and after artificial time misalignment compensation for constant attenuation factors.

With the results, it is apparent that the SSNR improvement decreased rapidly with an artificial time misalignment; however, the simple cross-correlation method had good performance that was independent of initial time misalignment. In particular, the SSNR improvement averaged less than 2 dB degradation compared to perfect time alignment. After the cross-correlation compensation, the average remaining misalignment was 0.2 samples (−10 dB), 0 samples (0 dB), and 0 samples (+10 dB).

7.4. Attenuation factor estimation

Fig. 6 demonstrates the effects of the artificial error, as measured by MSE, added to attenuation factors of a 24 microphone channel configuration for the multichannel LSA estimator with multichannel spectral phase estimator across white noise at 0 dB input SNR with constant attenuation factors.

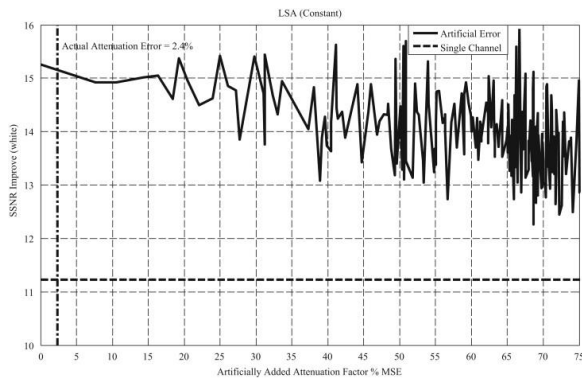


Fig. 6. SSNR improvement for multichannel LSA with multichannel spectral phase estimator due to artificial error in attenuation factors for constant attenuation factors.

Through the results, the misestimation of the constant attenuation factors due to artificially added error only causes a relatively small decrease of 1.3 dB in SSNR improvement. The actual computed error for the constant attenuation case was 31.5% (−10 dB input SNR), 2.4% (0 dB input SNR), and 0.31% (10 dB input SNR), which produces approximately 1–2 dB degradation. Theoretically, the worst-case impact for attenuation factor misestimation would occur when a single microphone channel has a dominatingly large attenuation factor, which reduces performance to the single channel estimator applied to that particular channel. Clearly, there is little SSNR improvement decrease from attenuation factor misestimation.

8. Conclusion

In this paper, optimal MMSE estimators were developed and implemented for STSA, LSA, and spectral phase for distributed microphone speech enhancement. The focus was to generalize the existing single channel STSA, LSA, and spectral phase estimators and present the benefits of utilizing the additional microphone information for providing gains in noise reduction, speech distortion, and overall speech quality, which were measured by the SSNR, LLR, and PESQ as objective metrics. From the experimental results, the multichannel LSA estimator showed significant improvements in SSNR improvement, LLR output, and PESQ output across several different noises and input SNRs, particularly when coupled with the multichannel spectral phase estimator, and demonstrated robustness due to time alignment and attenuation factor estimation. For future work, the multichannel STSA and LSA estimator cost functions along with speech prior and noise likelihood statistical models could be modified to obtain further improvements.

Acknowledgments

The authors would like to thank the National Science Foundation (Grant no. [IIS-0326395](#)) and U.S. Department of Education (GAANN Grant [P200A010104](#)) for supporting this work and Thomas Lotter and Christian Benien for providing invaluable insights into their multichannel speech enhancement research.

Appendix A

In this appendix, the MMSE STSA estimator is derived for distributed multichannel signals. After substitution of the statistical models of [\(4\)](#), [\(6\)](#), the result from [\(7\)](#) is

$$(25) \hat{A}_{STSA} = \frac{\int_0^\infty A^2 \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \exp\left(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2)/\sigma_{N_i}^2\right) d\alpha dA}{\int_0^\infty A \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \exp\left(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2)/\sigma_{N_i}^2\right) d\alpha dA}.$$

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

$$(26) \int_0^{2\pi} \exp\left(-\sum_{i=1}^M \frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right) d\alpha = \exp\left(-\sum_{i=1}^M \frac{|Y_i|^2 + c_i^2 A^2}{\sigma_{N_i}^2}\right) \int_0^{2\pi} \exp(a \cos \alpha + b \sin \alpha) d\alpha,$$

where

$$(27) a = \sum_{i=1}^M \frac{2c_i A}{\sigma_{N_i}^2} \text{Re}(Y_i),$$

and

$$(28) b = \sum_{i=1}^M \frac{2c_i A}{\sigma_{N_i}^2} \text{Im}(Y_i).$$

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

$$(29) a \cos \alpha + b \sin \alpha = \sqrt{a^2 + b^2} \cos(\alpha - \arctan(b/a)),$$

where

$$(30) \sqrt{a^2 + b^2} = 2A \left| \sum_{i=1}^M \frac{c_i Y_i}{\sigma_{N_i}^2} \right|.$$

Since the integral in (26) for the spectral phase α is over one full period, the spectral phase shift of $\arctan(b/a)$ is removed from (29). By means of equation 8.431.1 in [15], the integral in (26) is rewritten as

$$(31) \int_0^{2\pi} \exp(a \cos \alpha + b \sin \alpha) d\alpha = 2\pi I_0 \left(2A \left| \sum_{i=1}^M \frac{c_i Y_i}{\sigma_{N_i}^2} \right| \right),$$

which reduces (25) to

$$(32) \hat{A}_{STSA} = \frac{\int_0^\infty A^2 \exp(-A^2(1/\lambda)) I_0(2A \left| \sum_{i=1}^M \left(\frac{c_i Y_i}{\sigma_{N_i}^2} \right) \right|) dA}{\int_0^\infty A \exp\left(-A^2\left(\frac{1}{\lambda}\right)\right) I_0\left(2A \left| \sum_{i=1}^M \left(\frac{c_i Y_i}{\sigma_{N_i}^2} \right) \right| \right) dA}.$$

Through substitution of equations 8.406.3 and 6.631.1 in [22] and [15], (32) is written as

$$(33) \hat{A}_{STSA} = \frac{\Gamma(1.5) {}_1F_1\left(\left(\frac{3}{2}\right); 1; v\right)}{\left(\frac{1}{\lambda}\right)^{1/2} {}_1F_1(1; 1; v)},$$

where v is defined as

$$(34) v = \frac{\left| \sum_{i=1}^M (c_i Y_i / \sigma_{N_i}^2) \right|^2}{1/\lambda},$$

and $1/\lambda$ is given in (14) with ${}_1F_1(\cdot; \cdot; \cdot)$ denoted as the confluent hypergeometric function as described by equation 9.210 in [15]. Since the spectral amplitude A and spectral variance σ_S^2 are attenuated at each microphone i , giving $A_i = c_i A$ and $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$, (34) simplifies to (9), which is the multichannel extension of v given in [9]. From the relationship given by equation 9.212.1 in [15], (33) is rewritten as

$$(35) \hat{A}_{STSA} = \Gamma(1.5) \left(\frac{\sigma_S^2}{1 + \sum_{i=1}^M \xi_i} \right) {}_1F_1\left(-\frac{1}{2}; 1; -v\right).$$

For a more efficient implementation of the estimator, the confluent hypergeometric function in (35) can be replaced by the modified Bessel functions of the first kind of the 0th order and 1st order given by equations A.1.31a in [23] to produce the final closed-form solution of the multichannel STSA estimator \hat{A}_{STSA} given in (8).

Appendix B

In this appendix, the MMSE LSA estimator is derived for distributed multichannel signals. After substitution of (4), (6), (12) is expressed as

$$(36) \Phi_{Z|Y_1, \dots, Y_M}(\mu) = \frac{\int_0^\infty A^{\mu+1} \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \exp(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2 / \sigma_{N_i}^2)) d\alpha dA}{\int_0^\infty A \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \exp(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2 / \sigma_{N_i}^2)) d\alpha dA}.$$

The integration over the spectral phase α is performed exactly as in [Appendix A](#). By employing (26), (27), (28), (29), (30), (31) from [Appendix A](#), (36) is written as

$$(37) \Phi_{Z|Y_1, \dots, Y_M}(\mu) = \frac{\int_0^\infty A^{\mu+1} \exp(-A^2(1/\lambda)) I_0(2A |\sum_{i=1}^M (c_i Y_i / \sigma_{N_i}^2)|) dA}{\int_0^\infty A \exp(-A^2(1/\lambda)) I_0(2A |\sum_{i=1}^M (c_i Y_i / \sigma_{N_i}^2)|) dA}.$$

Through application of equations 8.406.3 and 6.631.1 in [\[22\]](#) and [\[15\]](#), the closed-form solution for $\Phi_{Z|Y_1, \dots, Y_M}(\mu)$ is established in [\[13\]](#).

The differentiation of [\[13\]](#) with respect to μ results in three derivative terms that are written as

$$(38) E[Z|Y_1, \dots, Y_M] = \frac{d}{d\mu} [\Phi_{Z|Y_1, \dots, Y_M}(\mu)]|_{\mu=0} = [\frac{d}{d\mu} (\Gamma(\frac{\mu}{2} + 1)) \frac{1}{(1/\lambda)^{\mu/2}} {}_1F_1(-\frac{\mu}{2}; 1; -v)]|_{\mu=0} + [\Gamma(\frac{\mu}{2} + 1) \frac{d}{d\mu} (\frac{1}{(1/\lambda)^{\mu/2}}) {}_1F_1(-\frac{\mu}{2}; 1; -v)]|_{\mu=0} + [\Gamma(\frac{\mu}{2} + 1) \frac{1}{(1/\lambda)^{\mu/2}} \frac{d}{d\mu} ({}_1F_1(-\frac{\mu}{2}; 1; -v))]|_{\mu=0},$$

and evaluated at $\mu = 0$. The derivative of the first term is evaluated exactly as in [\[10\]](#) using

$$(39) \frac{d}{d\mu} \left(\Gamma\left(\frac{\mu}{2} + 1\right) \right) = \Gamma\left(\frac{\mu}{2} + 1\right) \frac{d}{d\mu} \left(\ln \left(\Gamma\left(\frac{\mu}{2} + 1\right) \right) \right).$$

Through the series expansion given by equation 8.342.1 in [\[22\]](#), the last term in [\(39\)](#) is rewritten as

$$(40) \ln \left(\Gamma\left(\frac{\mu}{2} + 1\right) \right) = -c \frac{\mu}{2} + \sum_{r=2}^{\infty} \frac{(-\mu)^r}{2^r r} \alpha_r,$$

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

$$(41) \alpha_r \triangleq \sum_{n=1}^{\infty} \frac{1}{n^r}.$$

By differentiating [\(40\)](#) term-by-term and evaluating [\(39\)](#) at $\mu = 0$, the derivative of the first term in [\(38\)](#) is given as

$$(42) \frac{d}{d\mu} [\Gamma(\frac{\mu}{2} + 1)]|_{\mu=0} = \frac{c}{2}.$$

The derivative of the second term $(1/\lambda)^{-\mu/2}$ in (38) is computed in a straightforward manner by rewriting it in exponential form and evaluating at $\mu = 0$ as

$$(43) \frac{d}{d\mu} \left[\frac{1}{(1/\lambda)^{\mu/2}} \right] \Big|_{\mu=0} = \frac{d}{d\mu} [e^{(1/2)\mu \ln(\lambda)}] \Big|_{\mu=0} = \frac{1}{2} \ln(\lambda).$$

For the computation of the third term, the confluent hypergeometric function ${}_1F_1(-\mu/2; 1; -v)$ is differentiated through its series expansion from equation 9.210.1 in [15] as

$$(44) {}_1F_1(a; b; x) = \sum_{r=0}^{\infty} \frac{(a)_r x^r}{(c)_r r!},$$

where $(a)_r = 1a(a+1) \cdots (a+r-1)$ with $(a)_0 \triangleq 1$. By differentiating (44) term-by-term and evaluating at $\mu = 0$, the derivative is given as

$$(45) \frac{d}{d\mu} [{}_1F_1(-\frac{\mu}{2}; 1; -v)] \Big|_{\mu=0} = -\frac{1}{2} \sum_{r=1}^{\infty} \frac{(-v)^r}{r!} \frac{1}{r}.$$

By combining the three derivative results of (42), (43), and (45), (38) reduces to

$$(46) E[Z|Y_1, \dots, Y_M] = \frac{d}{d\mu} [\Phi_{Z|Y_1, \dots, Y_M}(\mu)] \Big|_{\mu=0} = (-\frac{c}{2}) {}_1F_1(0; 1; -v) + \ln(\sqrt{\lambda}) {}_1F_1(0; 1; -v) + (-\frac{1}{2} \sum_{r=1}^{\infty} \frac{(-v)^r}{r!} \frac{1}{r}) = -\frac{1}{2} [c + \sum_{r=1}^{\infty} \frac{(-v)^r}{r!} \frac{1}{r}] + \frac{1}{2} \ln(\lambda),$$

where ${}_1F_1(0; 1; -v) = 1$. From equations 8.211.1 and 8.214.1 in [22], (46) is rewritten as

$$(47) E[Z|Y_1, \dots, Y_M] = -\frac{1}{2} [-\int_v^{\infty} \frac{e^{-t}}{t} dt - \ln(v)] + \frac{1}{2} \ln(\lambda) = -\frac{1}{2} \ln(1/\lambda) + \frac{1}{2} [\int_v^{\infty} \frac{e^{-t}}{t} dt + \ln(v)].$$

After exponentiation of (47), the multichannel LSA estimator \hat{A}_{LSA} is written in (15) as

$$(48) \hat{A}_{LSA} = (\frac{v}{1/\lambda})^{1/2} \exp(\frac{1}{2} \int_v^{\infty} \frac{e^{-t}}{t} dt).$$

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 $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$, the first term in (15) can be rewritten as

$$(49) \left(\frac{v}{1/\lambda} \right)^{1/2} = \left(\frac{\sum_{i=1}^M (\xi_i / \gamma_i)}{\sum_{i=1}^M (c_i^2 / R_i^2)} \right)^{1/2} \left(\frac{\left| \sum_{i=1}^M (\sqrt{\xi_i} / \sigma_{N_i}) Y_i \right|}{1 + \sum_{i=1}^M \xi_i} \right).$$

The final closed-form solution of the multichannel LSA estimator is

$$(50) \hat{A}_{LSA} = \left(\frac{\sum_{i=1}^M (\xi_i / \gamma_i)}{\sum_{i=1}^M (c_i^2 / R_i^2)} \right)^{1/2} \left(\frac{\left| \sum_{i=1}^M (\sqrt{\xi_i} / \sigma_{N_i}) Y_i \right|}{1 + \sum_{i=1}^M \xi_i} \right) \exp \left(\frac{1}{2} \int_v^{\infty} \frac{e^{-t}}{t} dt \right).$$

Appendix C

In this appendix, the MMSE 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 $\{Y_1, \dots, Y_M\}$, (17) is rewritten as

$$(51) \min_{g, \rho} E[|e^{j\alpha} - g|^2 | Y_1, \dots, Y_M] + \rho(|g| - 1) = \min_{g, \rho} E[|\cos\alpha - g_R|^2 | Y_1, \dots, Y_M] + E[|\sin\alpha - g_I|^2 | Y_1, \dots, Y_M] + \rho(g_R^2 + g_I^2)^{1/2} - \rho,$$

which requires computation of the partial derivatives $\partial(E[\cdot])/\partial g_R = 0$, and $\partial(E[\cdot])/\partial g_I = 0$. The partial derivatives with respect to g_R and g_I are computed to find the solutions of $\partial(E[\cdot])/\partial g_R = 0$ and $\partial(E[\cdot])/\partial g_I = 0$ as

$$(52) g_R(2 + \rho) = 2E[\cos\alpha | Y_1, \dots, Y_M],$$

and

$$(53) g_I(2 + \rho) = 2E[\sin\alpha | Y_1, \dots, Y_M].$$

The fundamental relationship between the real and imaginary components is given in (20) with

$$(54) E[\cos\alpha | Y_1, \dots, Y_M] = \frac{\int_0^\infty \int_0^{2\pi} \cos\alpha p(Y_1, \dots, Y_M | A, \alpha) p(A, \alpha) d\alpha dA}{\int_0^\infty \int_0^{2\pi} p(Y_1, \dots, Y_M | A, \alpha) p(A, \alpha) d\alpha dA},$$

and

$$(55) E[\sin\alpha | Y_1, \dots, Y_M] = \frac{\int_0^\infty \int_0^{2\pi} \sin\alpha p(Y_1, \dots, Y_M | A, \alpha) p(A, \alpha) d\alpha dA}{\int_0^\infty \int_0^{2\pi} p(Y_1, \dots, Y_M | A, \alpha) p(A, \alpha) d\alpha dA},$$

which closely resemble the integration performed in (7), (10) but with different arguments in the expectation operators. After substituting the statistical models for the speech prior (4) and noise likelihood (6), (54), (55) are rewritten as

$$(56) E[\cos\alpha | Y_1, \dots, Y_M] = \frac{\int_0^\infty A \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \cos\alpha \exp(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2 / \sigma_{N_i}^2)) d\alpha dA}{\int_0^\infty A \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \exp(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2 / \sigma_{N_i}^2)) d\alpha dA},$$

and

$$(57) E[\sin\alpha | Y_1, \dots, Y_M] = \frac{\int_0^\infty A \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \sin\alpha \exp(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2 / \sigma_{N_i}^2)) d\alpha dA}{\int_0^\infty A \exp(-A^2/\sigma_S^2) \int_0^{2\pi} \exp(-\sum_{i=1}^M (|Y_i - c_i A e^{j\alpha}|^2 / \sigma_{N_i}^2)) d\alpha dA}.$$

By utilizing (26) from Appendix A, the inner integral over the spectral phase α in (56) is expanded as

$$(58) \int_0^{2\pi} \cos\alpha \exp\left(-\sum_{i=1}^M \frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right) d\alpha \propto \int_0^{2\pi} \cos\alpha \exp(ac\cos\alpha + bc\sin\alpha) d\alpha.$$

Through (29) from Appendix A, the integral over the spectral phase α in (58) is further rewritten as

$$(59) \int_0^{2\pi} \cos\alpha \exp(ac\cos\alpha + bc\sin\alpha) d\alpha = \int_0^{2\pi} \cos\alpha \cos(\alpha - \psi) d\alpha,$$

where

$$(60) \psi = \tan^{-1}(b/a),$$

and a , b , and $\sqrt{a^2 + b^2}$ are given in (27), (28), and (30), respectively, from Appendix A. Using the product-to-sum cosine trigonometric identity, (59) simplifies to

$$(61) \int_0^{2\pi} \cos\alpha \cos(\alpha - \psi) d\alpha = \frac{\sqrt{a^2 + b^2}}{2} \left[\cos(\psi) \int_0^{2\pi} d\alpha + \int_0^{2\pi} \cos(2\alpha - \psi) d\alpha \right] = \pi\sqrt{a^2 + b^2} \cos(\psi),$$

since the spectral phase shift of ψ in the second integral over the spectral phase α in (61) is irrelevant for the limits of integration. From (26) in Appendix A and (61), (58) is written as

$$(62) \int_0^{2\pi} \cos\alpha \exp\left(-\sum_{i=1}^M \frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right) d\alpha \propto \pi\sqrt{a^2 + b^2} \cos\psi.$$

In a similar manner, the inner integral over the spectral phase α in (57) is given by

$$(63) \int_0^{2\pi} \sin\alpha \exp\left(-\sum_{i=1}^M \frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right) d\alpha \propto \pi\sqrt{a^2 + b^2} \cos\theta,$$

where

$$(64) \theta = \sin^{-1}(a/\sqrt{a^2 + b^2}).$$

Through (62), (63), the expectations in (56), (57) are written as

$$(65) E[\cos\alpha | Y_1, \dots, Y_M] = \frac{\sqrt{a^2 + b^2}}{2} \cos\psi \frac{\int_0^\infty A \exp(-A^2(1/\lambda)) dA}{\int_0^\infty A \exp(-A^2(1/\lambda)) I_0(2A|\sum_{i=1}^M (c_i Y_i / \sigma_{N_i}^2)|) dA},$$

and

$$(66) E[\sin\alpha|Y_1, \dots, Y_M] = \frac{\sqrt{a^2+b^2}}{2} \cos\theta \frac{\int_0^\infty A \exp(-A^2(1/\lambda)) dA}{\int_0^\infty A \exp(-A^2(1/\lambda)) I_0(2A|\sum_{i=1}^M (c_i Y_i / \sigma_{N_i}^2)|) dA},$$

with $1/\lambda$ given by (14). By utilizing the expectations from (65), (66) and employing the definitions (60), (64), the multichannel spectral phase $\hat{\alpha}$ estimator from (21) is written as

$$(67) \hat{\alpha} = \tan^{-1}(\cos\theta/\cos\psi) = \tan^{-1}(b/a),$$

where a and b are specified in (27), (28) from Appendix A. Through simplification of the ratio 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 the multichannel spectral phase estimator $\hat{\alpha}$ in (67) is given in (21).

References

- [1] P.E. Papamichalis. **Practical Approaches to Speech Coding**. Prentice-Hall, New York, NY (1978)
- [2] S.R. Quackenbush, I.T.P. Barnwell, M.A. Clements. **Objective Measures of Speech Quality**. Prentice-Hall, New York (1998)
- [3] ITU, Perceptual Evaluation of Speech Quality (PESQ), and Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs., ITU-T Recommendation, 2001.
- [4] Y. Hu, P. Loizou. **Evaluation of objective quality measures for speech enhancement**. IEEE Transactions on Audio, Speech, and Language Processing, 16 (2008), pp. 229-238
- [5] J. Polastre, R. Szwedczyk, A. Mainwaring. Chapter 18: Analysis of Wireless Sensor Networks for Habitat Monitoring, in Wireless Sensor Networks, Kluwer Academic Publishers, Norwell, MA (2004)
- [6] B. Widrow, J.R. Glover Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong Jr., R.C. Goodlin. **Adaptive noise cancelling: principles and applications**. Proceedings of the IEEE, 63 (1975), pp. 1692-1716
- [7] M. Brandstein, D. Ward. **Microphone Arrays**. Springer-Verlag, New York, NY (2001)
- [8] I.A. McCowan. **Robust Speech Recognition using Microphone Arrays**. Queensland University of Technology (2001)
- [9] Y. Ephraim, D. Malah. **Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator**. IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-32 (1984), pp. 1109-1121
- [10] Y. Ephraim, D. Malah. **Speech enhancement using a minimum mean-square error log-spectral amplitude estimator**. IEEE Transactions on Acoustics, Speech and Signal Processing, 33 (1985), pp. 443-445
- [11] T. Lotter, C. Benien, P. Vary. **Multichannel direction-independent speech enhancement using spectral amplitude estimation**. EURASIP Journal on Applied Signal Processing (2003), pp. 1147-1156
- [12] B.D.V. Veen, K.M. Buckley. **Beamforming: a versatile approach to spatial filtering**. IEEE ASSAP Magazine (1988)
- [13] H.L. v. Trees. Detection, Estimation, and Modulation Theory, vol. I, Wiley, New York, NY (1968)
- [14] H.V. Poor. An Introduction to Signal Detection and Estimation (2nd ed.) (1994)
- [15] I.S. Gradshteyn, Z.M. Ryzhik. **Table of Integrals, Series, and Products**. (5th ed.), Academic, New York (1994)
- [16] M.B. Trawicki, M.T. Johnson, Optimal Distributed Microphone Phase Estimation, presented at International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Taipei, Taiwan, ROC, 2009.
- [17] J. Garofolo, L. Lamel, W. Fisher. **TIMIT acoustic-phonetic continuous speech corpus**. Linguistic Data Consortium (1993)
- [18] A. Varga, 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, 12 (1993), pp. 247-251

- [19] O. Cappe. **Elimination of the musical noise phenomenon with the ephraim and malah noise suppression.** IEEE Transactions on Speech and Audio Processing, 2 (1994), pp. 345-349
- [20] C.H. Knapp, G.C. Carter. **The generalized correlation method for estimation of time delay.** IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-24 (1976), pp. 320-327
- [21] L.E. Kinsler. Fundamentals of Acoustics (4th ed.), John Wiley & Sons, Inc. (1999)
- [22] I.S. Gradshteyn, Z.M. Ryzhik. **Table of Integrals, Series, and Products.** Academic, New York City, New York (1980)
- [23] D. Middleton. **Introduction to Statistical Communication Theory.** McGraw-Hill, New York (1960)