

Beta-order minimum mean-square error multichannel spectral amplitude estimation for speech enhancement

M. B. Trawicki^{*,†} and M. T. Johnson

*Department of Electrical and Computer Engineering, Speech and Signal Processing Laboratory,
Marquette University Milwaukee, WI 53201-1881, USA*

In this paper, the minimum mean-square error (MMSE) β -order estimator for multichannel speech enhancement is proposed. The estimator is an extension of the single-channel MMSE β -order and multichannel MMSE short-time spectral amplitude estimators using Rayleigh and Gaussian distributions for the statistical models under the assumption of a diffuse noise field where the noise is estimated independently across each of the microphones. Experiments are performed to evaluate the new estimator against the baseline single-channel and multichannel estimators using various values of the β parameter and number of microphones along with different levels of noises as a function of the input signal-to-noise ratio. By the utilization of additional microphones, the multichannel MMSE β -order estimator achieves performance gains in noise reduction, speech distortion, and speech quality as measured by the segmental signal-to-noise ratio, log-likelihood ratio, and perceptual evaluation of speech quality objective metrics. Copyright © 2015 John Wiley & Sons, Ltd.

Received 27 January 2013; Revised 12 September 2014; Accepted 15 December 2014

KEY WORDS: acoustic arrays; speech enhancement; 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 SNR (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]. Although 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, 8–12]. By comparison, there has been relatively little research for distributed microphone configurations [13] where the microphones are spread throughout a large area of interest with unknown spacing and geometry and array assumptions do not hold anymore.

*Correspondence to: Marek B. Trawicki, Electronic and Computer Engineering, Marquette University, Speech and Signal Processing Laboratory, Milwaukee, WI, USA.

†E-mail: marek.trawicki@marquette.edu

In order to advance the current state-of-the-art speech enhancement methods for distributed microphones [14–16], it is important to generalize the existing work from single-channel microphones, dual channel microphones, and microphone arrays.

In speech enhancement, there has been much work on single-channel estimation of the spectral amplitude. From the foundational work involving the minimum mean-square error (MMSE) estimation of the short-time spectral amplitude (STSA) [9] and log-spectral amplitude (LSA) [10], researchers have modified the STSA and LSA cost functions to achieve further improvements in noise reduction along with decreases in speech distortion and increases in speech quality. Specifically, the STSA cost function was generalized to the β -order cost function [17, 18], which incorporates a power law parameter β on the STSA cost function. As $\beta = 1$ and $\beta \rightarrow 0$, the MMSE β -order spectral amplitude estimator is equivalent to the MMSE STSA and LSA estimators [17]. From You et al. [17], the strong speech spectral amplitudes were attenuated by almost the same amount for high and low β values with high instantaneous SNR values. In contrast, the weak speech spectral amplitudes were primarily preserved because the gain value is big for the low instantaneous SNR spectral amplitudes with large β values. Based on the experimental results, the single-channel MMSE β -order estimator yielded a good trade-off between speech distortion and residual noise reduction, particularly for weak spectral components. As shown by Plourde and Champagne [18], the negative values of β introduced more speech distortion but produced more noise reduction. The normalization in the β -order cost function with $\beta < 0$ penalized the estimation error more heavily for spectral valleys rather than spectral peaks. Therefore, the single-channel MMSE β -order estimator provided a better overall estimation of the speech in the spectral valleys.

In the single-channel MMSE STSA, LSA, and β -order estimators, Rayleigh and Gaussian probability density functions (PDFs) were utilized as the standard statistical models; however, the distributions have been recently improved to more accurately model the joint PDF of the speech spectral amplitude and spectral phase along with the conditional PDF of the observed noisy spectral amplitude given the spectral amplitude and spectral phase. Martin [19] presented MMSE estimators of the discrete Fourier transform coefficients rather than the spectral amplitude and utilized super-Gaussian models for the speech and noise components, namely Gaussian, Laplace, and Gamma distributions. Lotter and Vary [20] integrated the same Laplace and Gamma super-Gaussian speech priors into maximum *a posteriori* (MAP) estimators of the spectral amplitude, and Erkelens et al. [21] extended the MAP spectral amplitude estimators into MMSE spectral amplitude estimators using the generalized Gamma speech priors with Rayleigh noise PDFs. Andrianakis and White [22] continued with the MMSE and MAP spectral amplitude estimators using Gamma distribution but introduced Chi distribution for modeling the speech priors, and Breithaupt et al. [23] developed a MMSE STSA estimator using a variable compression function in the error criterion and the utilized the Chi distribution as the speech prior. From the incorporation of super-Gaussian statistical models, the squared-error cost functions demonstrated improvements over the Rayleigh speech prior distributions. In the derivation of the multichannel MMSE β -order estimator, the distributions were selected for their ability to accurately fit the models and facilitate derivation of the statistical estimators.

Despite the success of the single-channel MMSE β -order estimator, there is not a MMSE β -order estimator for multichannel enhancement. The framework exists for the multichannel MMSE STSA [11] and multichannel MMSE LSA [15] estimators, which have demonstrated performance improvements over their single-channel MMSE STSA and LSA estimator counterparts. For the multichannel MMSE STSA estimator, the spectral amplitude of the source signal was estimated at each of the microphones but rewritten to estimate the spectral amplitude of the true source signal at an arbitrary reference microphone. The multichannel STSA and LSA estimators incorporated the multichannel spectral phase estimator that utilized local information from the individual microphones to determine an estimate of the spectral phase and improved the noise reduction performance. In these multichannel speech enhancement estimators, Rayleigh and Gaussian distributions were used for the speech and noise PDFs while the noise field was assumed to be a diffuse noise field. Hendriks *et al.* [24] expanded the multichannel STSA and LSA estimators by integrating a generalized Gamma speech prior distribution and noise correlation matrix, which did not assume a diffuse noise field and accounted for all the available acoustic and spatial information of the speech

and noise sources in the environment. The subsequent multichannel MMSE estimator is rather limited by its assumption of $\beta = 1$ because it is only a special case of the single-channel MMSE β -order estimator. To relax this restriction, the multichannel MMSE β -order estimator derived in this work will assume $\beta \neq 1$ to fully exploit the efficacy of the β -order cost function and Rayleigh and Gaussian distributions for the speech and noise statistical models and diffuse noise field, where the noises are estimated independently across each of the microphones because the magnitude-squared coherence (MSC) $C_{ij}(f) = \text{sinc}(2\pi d_{ij}/c)$, where d_{ij} is the distance d between two microphones i and j , f is the frequency, and c is the speed of sound, is approximately small for high frequencies outside the primary energy of speech. Generally, 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 goal is to extend the single-channel MMSE STSA β -order cost estimator for achieving additional gains in multichannel speech enhancement covering large areas of interest and providing a theoretical foundation for incorporating additional acoustical effects into the model.

The remainder of this paper is organized into the following sections: multichannel system (Section 2), beta-order cost function (Section 3), experimental methodology (Section 4), experimental results (Section 5), and conclusion (Section 6).

2. MULTICHANNEL SYSTEM

Consider an arbitrary array of M microphones. At each microphone i , the static source signal $s(t)$ is captured as time-delayed and attenuated coherent clean signals $c_i s(t - \tau_i)$ corrupted by additive and uncorrelated noise $n_i(t)$ between the different microphones with time-invariant attenuation factors c_i and time-delays τ_i . Without loss of generality, the first microphone, $i = 1$, is assumed as the reference microphone with $c_1 = 1$. The propagation model in the time-domain is given as

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

where the time-delays τ_i have been removed from (1) after accurately performing time-aligned through cross-correlation methods. The frequency domain representation of (1) is expressed as

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

where l and k represent the frame and frequency bin with noisy and clean spectral amplitudes R_i and A , noisy and clean spectral phases ϑ_i and α , and spectral noise N_i for each individual microphone i . To simplify the notation for the upcoming sections, the variables in (2) will subsequently be written without the explicit dependencies on l and k .

3. BETA-ORDER COST FUNCTION

Based on the STSA and LSA cost functions, the β -order cost function is given as

$$d(A, \hat{A}; \beta) = (A^\beta - \hat{A}^\beta)^2 \quad (3)$$

for some constant β parameter. In a similar fashion to the multichannel MMSE STSA ($\beta = 1$) [11] and MMSE LSA ($\beta = 0.001$) [15] estimators, the minimization of Bayes risk using the cost function in (3) with respect to \hat{A} results in the proposed multichannel MMSE β -order estimator

$$\hat{A}^\beta = \int_0^\infty \int_0^{2\pi} A^\beta p(Y_1, \dots, Y_M | A, \alpha) p(A, \alpha) d\alpha dA \bigg/ \int_0^\infty \int_0^{2\pi} p(Y_1, \dots, Y_M | A, \alpha) p(A, \alpha) d\alpha dA. \quad (4)$$

3.1. Statistical models

Rayleigh distributions are assumed for the joint speech spectral amplitude and spectral phase PDF

$$p(A, \alpha) = (A/\pi\sigma_s^2) \exp(-A^2/\sigma_s^2), \quad (5)$$

and Gaussian distributions are assumed for the conditional noise given spectral amplitude and spectral phase PDF

$$p(Y_i | A, \alpha) = \left(1/\pi\sigma_{N_i}^2\right) \exp\left(-|Y_i - c_i A e^{j\alpha}|^2/\sigma_{N_i}^2\right), \quad (6)$$

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, the correlation of the noises between the different microphones is approximately low for high frequencies with relatively large microphone distances according to the MSC (i.e., $C_{ij}(f) < 0.1$ with $d_{ij} > \sim 14$ cm). Therefore, the noises are uncorrelated at each of the microphones, which results in the conditional joint distribution of the noisy spectral observations $\{Y_1, \dots, Y_M\}$ given the spectral amplitude and spectral phase written as

$$p(Y_1, \dots, Y_M | A, \alpha) = \prod_{i=1}^M p(Y_i | A, \alpha) = \prod_{i=1}^M \left(1/\pi\sigma_{N_i}^2\right) \exp\left(-\sum_{i=1}^M |Y_i - c_i A e^{j\alpha}|^2/\sigma_{N_i}^2\right). \quad (7)$$

3.2. Optimal estimator

From the statistical models (5) and (7), the multichannel MMSE β -order estimator in (4) is

$$\begin{aligned} \hat{A}^\beta &= \int_0^\infty A^{\beta+1} \exp\left(-\frac{A^2}{\sigma_s^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^M \frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right) d\alpha dA \\ &\quad / \int_0^\infty A \exp\left(-\frac{A^2}{\sigma_s^2}\right) \int_0^{2\pi} \exp\left(-\sum_{i=1}^M \frac{|Y_i - c_i A e^{j\alpha}|^2}{\sigma_{N_i}^2}\right) d\alpha dA. \end{aligned} \quad (8)$$

As in [15], the spectral phase α is integrated out from the inner integrals

$$\begin{aligned} \hat{A}^\beta &= \int_0^\infty A^{\beta+1} \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A \left|\sum_{i=1}^M \frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right) dA \\ &\quad / \int_0^\infty A \exp\left(-A^2 \frac{1}{\lambda}\right) I_0\left(2A \left|\sum_{i=1}^M \frac{c_i Y_i}{\sigma_{N_i}^2}\right|\right) dA, \end{aligned} \quad (9)$$

where $I_0(\bullet)$ denotes the modified Bessel function of the first kind of the 0th-order and

$$1/\lambda = 1/\sigma_s^2 + \sum_{i=1}^M c_i^2/\sigma_{N_i}^2. \quad (10)$$

By utilizing 8.406.3 and 6.631.1 in [25], the closed-form solutions for (9) is given in terms of the gamma function $\Gamma(\bullet)$

$$\Gamma(h) = \int_0^\infty t^{h-1} e^{-t} dt \quad (11)$$

and confluent hypergeometric function ${}_1F_1(\bullet; \bullet; \bullet)$ (described by 9.210.1 in [25])

$${}_1F_1(a; b; c) = 1 + \frac{a}{b} \frac{c}{1!} + \frac{a(a+1)}{b(b+1)} \frac{c^2}{2!} + \frac{a(a+1)(a+2)}{b(b+1)(b+2)} \frac{c^3}{3!} + \dots \quad (12)$$

as

$$\hat{A}^\beta = \left(\Gamma(\beta/2 + 1) / (1/\lambda)^{\beta/2} \right) ({}_1F_1((\beta + 2)/2; 1; z) / {}_1F_1(1; 1; z)), \quad (13)$$

where

$$z = \left| \sum_{i=1}^M \sqrt{\xi_i \gamma_i} e^{j\vartheta_i} \right|^2 / \left(1 + \sum_{i=1}^M \xi_i \right) \quad (14)$$

with *a priori* $\xi_i = \sigma_{S_i}^2 / \sigma_{N_i}^2$ and *a posteriori* $\gamma_i = R_i^2 / \sigma_{N_i}^2$ SNRs at each of the corresponding microphones i . From the relationship given by Equation 9.212.1 in [25], (13) is rewritten as

$$\hat{A}^\beta = \Gamma(\beta/2 + 1) \left(\sigma_S^2 / \left(1 + \sum_{i=1}^M \xi_i \right) \right)^{\frac{\beta}{2}} {}_1F_1(-\beta/2; 1; -z). \quad (15)$$

By taking the β^{th} root of both sides of (15), the final closed-form solution

$$\hat{A} = [\Gamma(\beta/2 + 1) {}_1F_1(-\beta/2; 1; -z)]^{\frac{1}{\beta}} \left(\sigma_S^2 / \left(1 + \sum_{i=1}^M \xi_i \right) \right)^{\frac{1}{2}}, \quad (16)$$

where $\beta > -2$. For the case of $M = 1$, (16) simplifies to the single-channel MMSE β -order estimator given as

$$\hat{A} = [\Gamma(\beta/2 + 1) {}_1F_1(-\beta/2; 1; -z)]^{\frac{1}{\beta}} (\sigma_S^2 / (1 + \xi))^{\frac{1}{2}}, \quad (17)$$

where

$$z = \xi \gamma / (1 + \xi). \quad (18)$$

4. EXPERIMENTAL METHODOLOGY

The proposed optimal multichannel MMSE β -order estimator derived in (16) was evaluated in MATLAB by simulating multiple microphone noisy signals with the TIMIT [26] and NOISEX [27] corpora. Specifically, the simulated noisy signals, which averaged 2.4 ± 0.5 s, were sampled at 16 kHz and corrupted by white, pink, and babble noises and created according to (1) with equal number of uncorrelated noises as microphones for 1 to 10 microphones. Although the signals were assumed to be perfectly synchronized without any time misalignment, previous work has illustrated that cross-correlation methods can accurately estimate time delays in the signals and effectively time align signals without any significant degradation in the enhancement results [15]. To demonstrate the best-case results, constant attenuation factors ($c_i = 1$), which represent the equal amplitude reduction between the original acoustic clean source signal and recorded noisy signals, were estimated at each of the microphones using the signal powers of the noisy signals across an entire utterance [15]. At each of the non-reference microphones, the noises were scaled according to the noise at the reference microphone and added to each of the attenuated clean signals at various input SNR levels. The noisy signals were truncated to produce an equal number of samples in each frame. Analysis conditions consisted of frames of 256 samples (25.6 ms) with 50% overlap using Hanning windows. Noise estimation was performed on the initial five frames of silence. The decision-directed [9] smoothing approach was utilized to estimate ξ_i with $\alpha_{SNR} = 0.98$ using thresholds of $\xi_{\min} = 10^{-25/10}$ for perceptual reasons [28] and $\gamma_{\min} = 40$ (implemented as a floor on $\sigma_{N_i}^2$) determined empirically to avoid numeric overflows and the spectral phase α was estimated as a quotient of two weighted sum of noisy spectral observations [15]. Objective measures of SSNR, LLR, and PESQ [4] were utilized to measure the noise reduction, speech distortion, and overall quality averaged over 10 enhanced signals that were reconstructed using the overlap-add technique.

5. EXPERIMENTAL RESULTS

With input SNRs of -10 dB to $+10$ dB at increments of $+10$ dB, the average (across β and noises) input LLR and input PESQ were 1.210, 0.995, and 0.6805 and 1.3704, 1.9864, and 2.6966. The baseline methods were the single-channel MMSE STSA ($\beta = 1$), LSA ($\beta = 0.001$), and β -order

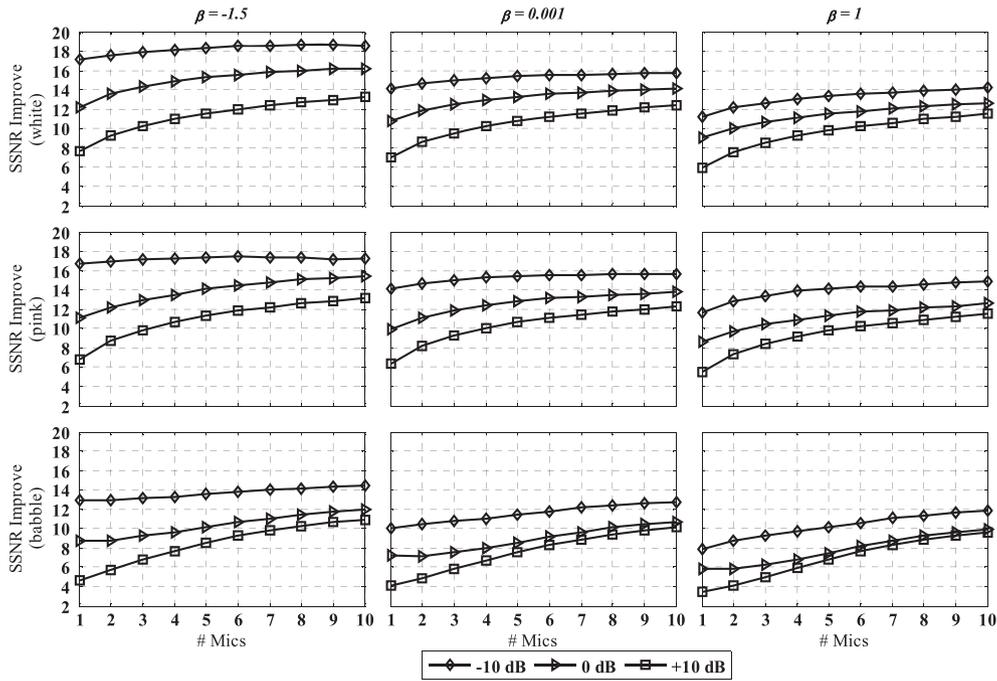


Figure 1. Signal-to-noise ratio improvement.

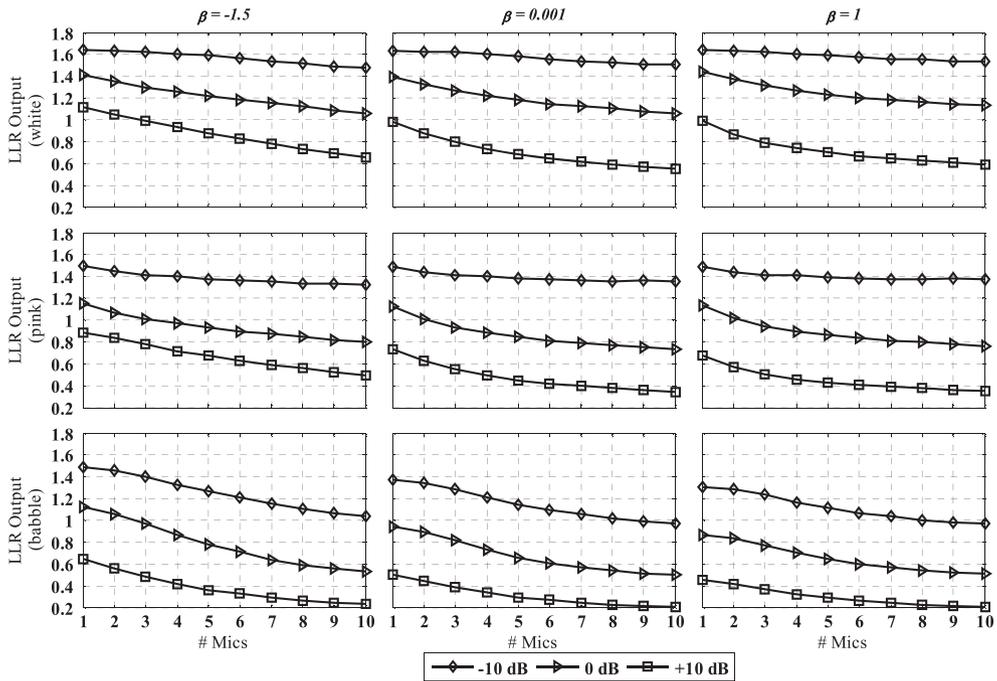


Figure 2. Log-likelihood ratio output.

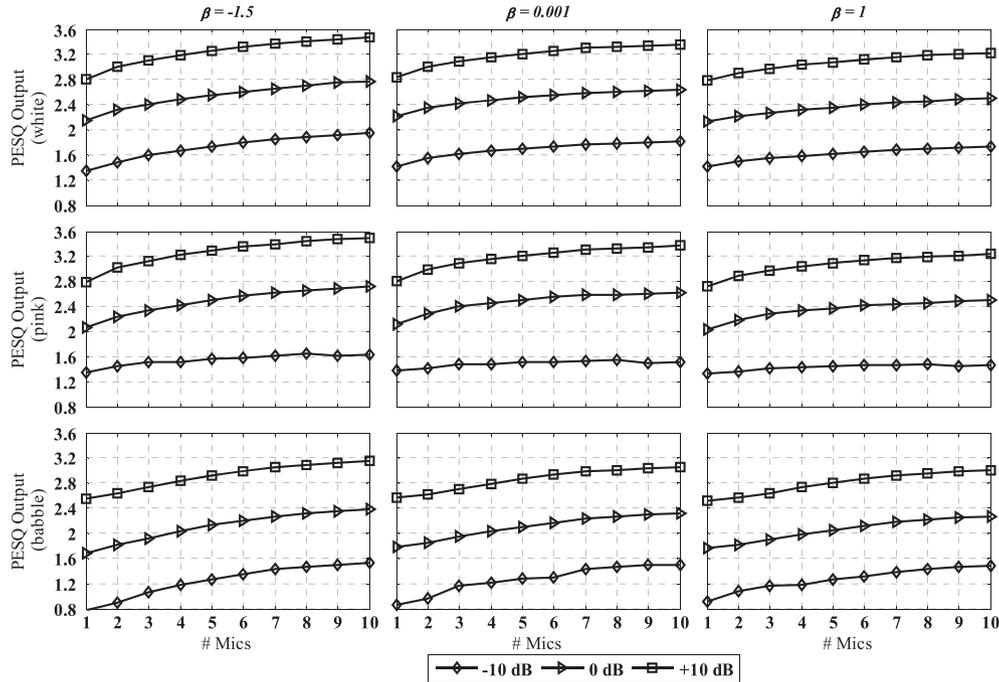


Figure 3. Perceptual evaluation of speech quality output.

estimators along with the multichannel MMSE STSA and LSA estimators. Figures 1, 2, and 3 show the SSNR improvement, LLR output, and PESQ output as a function of the number of microphones in the array across the noises.

Based on the results, the multichannel MMSE β -order estimator demonstrated gains over the single-channel MMSE STSA, LSA, and β -order estimators across the noises. In terms of SSNR improvements, the multichannel MMSE β -order estimator provided 2–4 dB increase in noise reduction over the baseline single-channel β -order estimator. In fact, there was a 1 dB decrease for each subsequent increase in the value of the β parameter towards positive values. Consequently, the multichannel MMSE β -order estimator illustrated a 2–3 dB gain over the multichannel MMSE STSA and LSA estimators for 10 microphones. For the LLR outputs, there were decreases in speech distortion from one microphone to 10 microphones essentially independent of the noises and β -order parameters along with single-channel and multichannel estimators. At the noisiest (–10 dB and 0 dB) and cleanest (+10 dB) input SNRs, the LLR outputs were 0.1 and 0.3 along with 0.7 from one microphone to 10 microphones. Although the single-channel and multichannel estimators yielded relatively similar LLR outputs across the noises and β -order parameters, the more negative β -order parameters ($\beta = -1.5$) had a much sharper decline in value than the more positive β -order parameters ($\beta = 1$). With the PESQ outputs, there was an increase in speech quality of 0.8 from one microphone to 10 microphones consistently across the noises and β parameters for each of the input SNRs. As the β parameter increased towards the multichannel MMSE STSA and LSA estimators for 10 microphones, the output PESQ decreased by 0.2–0.4 from the maximum of 2.0 (–10 dB) to 3.6 (+10 dB) using the multichannel MMSE β -order estimator. Overall, the multichannel MMSE β -order estimator generated improvements over the single-channel MMSE STSA, LSA, and β -order estimators with the recommendation of more negative β parameter ($\beta \rightarrow -2$).

6. CONCLUSION

The multichannel MMSE β -order estimator was derived for multichannel speech enhancement under the assumption of a diffuse noise field. In general, the majority of large area practical noisy

environments involve noise situations that are best characterized by a diffuse noise field, which allows for estimation of the noise statistics at each of the corresponding microphones. Because 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. By utilizing additional microphones, the focus of this research was to generalize the single-channel MMSE β -order for multichannel speech enhancement and demonstrate performance increases in noise reduction and speech quality and decreases in speech distortion. From the experimental results, the multichannel MMSE β -order estimator showed significant gains in SSNR improvements along with LLR and PESQ outputs over the baseline single-channel MMSE STSA, LSA, and β -order estimators and multichannel MMSE STSA and LSA estimators across different noises. For future work, the multichannel MMSE β -order estimator could offer further gains with modifications to the speech prior.

REFERENCES

1. Papamichalis PE. *Practical Approaches to Speech Coding*. Prentice-Hall: New York, NY, 1978.
2. Quackenbush SR, Barnwell I TP, Clements MA. *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. Hu Y, Loizou P. Evaluation of objective quality measures for speech enhancement. *IEEE Transactions on Audio, Speech, and Language Processing* 2008; **16**:229–238.
5. Polastre J, Szewczyk R, Mainwaring A. Chapter 18: Analysis of wireless sensor networks for habitat monitoring. In *Wireless Sensor Networks*, Vol. 5, Raghavendra C.S., Sivalingam K.M., Znati T. (eds). Kluwer Academic Publishers: Norwell, MA, 2004; 399–423.
6. Widrow B, Glover JR, Glover JR, Jr., Kaunitz J, Williams CS, Hearn RH, Zeidler JR, Dong, E, Jr., Goodlin RC. Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE* 1975; **63**:1692–1716.
7. Brandstein M, Ward D. *Microphone Arrays*. Springer-Verlag: New York, NY, 2001.
8. McCowan IA. *Robust Speech Recognition using Microphone Arrays*. Queensland University of Technology: Brisbane QLD 4001, Australia, 2001.
9. Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing* 1984; **ASSP-32**:1109–1121.
10. Ephraim Y, Malah D. Speech enhancement using a minimum mean-square error log-spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing* 1985; **33**:443–445.
11. Lotter T, Benien C, Vary P. Multichannel direction-independent speech enhancement using spectral amplitude estimation. *EURASIP Journal on Applied Signal Processing* 2003; **2003**:1147–1156.
12. Veen BDV, Buckley KM. Beamforming: A versatile approach to spatial filtering. *IEEE ASSAP Magazine* 1988; **5**:1147–1156.
13. Trawicki MB. Distributed multichannel processing for signal enhancement. In *Electrical and Computer Engineering*. Marquette University: Milwaukee, 2009; 1–228.
14. Trawicki MB, Johnson MT. Optimal distributed microphone phase estimation, presented at *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, R.O.C., 2009; 2177–2810.
15. Trawicki MB, Johnson MT. Distributed multichannel speech enhancement with minimum mean-square error short-time spectral amplitude, log-spectral amplitude, and spectral phase estimation. *Signal Processing* 2012; **92**: 345–356.
16. Trawicki M, Johnson MT. Distributed multichannel speech enhancement based on perceptually-motivated Bayesian estimators of the spectral amplitude. *IET Signal Processing* 2013; **7**:337–344.
17. You CH, Koh SN, Rahardja S. Beta-order MMSE spectral amplitude estimation for speech enhancement. *IEEE Transactions on Speech and Audio Processing* 2005; **13**:475–486.
18. Plourde E, Champagne B. Further analysis of the beta-order MMSE STSA estimator for speech enhancement, presented at *International Conference on Acoustics, Speech, and Signal Processing*, Vancouver, BC, 2007; 1594–1597.
19. Martin R. Speech enhancement based on minimum mean-square error estimation and supergaussian priors. *IEEE Transactions on Acoustics, Speech and Signal Processing* 2005; **13**:845–856.
20. Lotter T, Vary P. Speech enhancement by MAP spectral amplitude estimation using a Super-Gaussian speech model. *EURASIP Journal on Applied Signal Processing* 2005; **2005**:1110–1126.
21. Erkelens JS, Hendriks RC, Heusdens R, Jensen J. Minimum mean-square error estimation of discrete fourier coefficients with generalized gamma priors. *IEEE Transactions on Audio, Speech, and Language Processing* 2007; **15**:1741–1752.

22. Andrianakis I, White PR. Speech spectral amplitude estimators using optimally-shaped gamma and chi priors. *Speech Communication* 2009; **51**:1–14.
23. Breithaupt C, Krawczyk M, Martin R. Parameterized MMSE spectral magnitude estimation for the enhancement of noisy speech, presented at *International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, 2008; 4037–4040.
24. Hendriks RC, Heusdens R, Kjens U, Jensen J. On optimal multichannel mean-squared error estimators for speech enhancement. *IEEE Signal Processing Letters* 2009; **16**:885–888.
25. Gradshteyn IS, Ryzhik IM. *Tables of Integrals, Series, and Products*: Burlington, MA, 2007.
26. Garofolo J, Lamel L, Fisher W. TIMIT acoustic-phonetic continuous speech corpus. *Linguistic Data Consortium* 1993; Data.
27. Varga A, Steeneken HJM. 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* 1993; **12**: 247–251.
28. Cappe O. Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppression. *IEEE Transactions on Speech and Audio Processing* 1994; **2**:345–349.