International Journal of Theoretical and Applied Mathematics
2016; 2(2): 115-120
http://www.sciencepublishinggroup.com/j/ijtam
doi: 10.11648/j.ijtam.20160202.23

Multichannel MMSE Wiener Filter Using Complex Real and Imaginary Spectral Coefficients for Distributed Microphone Speech Enhancement

Marek B. Trawicki, Michael T. Johnson

Department of Electrical and Computer Engineering, Marquette University, Milwaukee, USA

Email address:
marek.trawicki@marquette.edu (M. B. Trawicki), mike.johnson@marquette.edu (M. T. Johnson)

To cite this article:
Marek B. Trawicki, Michael T. Johnson. Multichannel MMSE Wiener Filter Using Complex Real and Imaginary Spectral Coefficients for Distributed Microphone Speech Enhancement. *International Journal of Theoretical and Applied Mathematics.* Vol. 2, No. 2, 2016, pp. 115-120. doi: 10.11648/j.ijtam.20160202.23

Received: October 26, 2016; **Accepted**: November 26, 2016; **Published**: December 20, 2016

Abstract: In this paper, the authors propose a frequency domain multichannel Wiener filter for distributed microphone speech enhancement using acoustic arrays. The current state-of-the-art single channel estimators achieve noticeable performance gains using the to-noise ratio (SNR) and segmental signal-to-noise ratio (SSNR) objective measures, which measure noise reduction, but only achieve marginal performance gains using the Log-Likelihood Ratio (LLR) and Perceptual Evaluation of Speech Quality (PESQ) objective metrics, which correlate better than SNR and SSNR with speech distortion and overall speech quality. By extending the traditional single channel Wiener filter to multiple distributed channels through minimum mean-square error (MMSE) estimation of the complex real and imaginary components, the approach presented here demonstrates increases in the SSNR, LLR, and PESQ objective measures. Experimental results show that the new multichannel Wiener filter using distributed microphones produces gains of 5.0 dB (SSNR improvement), 0.7 (LLR output), and 0.8 (PESQ output) averaged across the 0 dB, 5 dB, and 10 dB input SNRs over the baseline single channel Wiener filter.

Keywords: Acoustic Arrays, Speech Enhancement, Parameter Estimation

1. Introduction

Wiener filtering is an optimal and a traditional baseline method for performing speech enhancement in either the time-domain [1] or frequency-domain [2] on noisy signals, which was originally developed and implemented for single channel microphones. For single channel methods, the work typically concentrates on frequency domain statistical estimators derived in the minimum mean-square error (MMSE) sense for estimation of the spectral amplitude [3-5] or complex real and imaginary components [2]. Through modifications to the statistical prior models or estimator equations, Andrianakis [6], Erkelens [7], Plourde [8], and You [9] demonstrated only marginal improvements over the corresponding baseline methods in the objective measures of Segmental Signal-to-Noise Ratio (SSNR), Log-Likelihood Ratio (LLR) and Perceptual Evaluation of Speech Quality (PESQ), which serve to predict noise reduction, speech distortion, and overall speech quality [10]. In order to achieve further gains in performance, these single channel estimators can be extended to multiple microphones [11, 12], particularly the distributed microphone paradigm [13].

Current multichannel Wiener filtering techniques can be categorized typically into methods using dual channel microphones [14] and microphone arrays [15-17]. The fundamental characteristics of those techniques depend on the assumptions of the microphone configurations. Whereas dual channel microphones require a reference noise microphone [14], microphone arrays [18] require close-spacing of the microphone elements and *a priori* knowledge of the geometry. For both of these microphone configurations, the noise is assumed to be correlated across the microphone channels. In contrast, these assumptions are no longer valid for distributed microphones since the microphones are widely dispersed to provide broad acoustic coverage over a given region and the ambient noise is incoherent across the channels [19]. The goal of this work is to develop and implement the multichannel Wiener filter for distributed microphones without the strict array assumptions [18] and illustrate improvements in SSNR, LLR, and PESQ with the additional microphone information.

The remainder of this paper is organized into the following sections: distributed microphone system (Section 0), multichannel Wiener filter (Section 0), simulation experiments and results (Section 0), and conclusion (Section 0).

2. Distributed Microphone System

Consider an arbitrary array of M microphones, where a particular microphone is represented as $i \in [1,...,M]$. At each microphone i, the unknown, omni-directional, spatially-stationary 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 $n_i(t)$ 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 $c_1=1$. Based on this distributed microphone scenario, the propagation model in the time-domain is given as

$$y_i(t) = c_i s(t) + n_i(t), \tag{1}$$

which can be accurately time-aligned through simple cross-correlation methods [20]. The frequency domain representation of (1) is expressed as

$$\begin{gathered} Y_i(\lambda,k) = c_i S(\lambda,k) + N_i(\lambda,k) \\ Y_{iR}(\lambda,k) + jY_{iI}(\lambda,k) = c_i\left[S_R(\lambda,k) + jS_I(\lambda,k)\right] + N_i(\lambda,k) \end{gathered}, \tag{2}$$

without the explicit dependencies on the frame λ and frequency bin k, where the noisy and clean real R and imaginary I spectral components are written in compact form as $Y_{i,(R,I)}$ and $S_{R,I}$.

3. Multichannel Wiener Filter

As a basic approach to speech enhancement, the multichannel Wiener filter is derived for distributed microphones as an extension to the single channel Wiener filter [2]. Through Bayes rule, the MMSE estimate of the real and imaginary spectral components of the clean spectral source $S_{R,I}$ is expressed as

$$\hat{S}_{R,I} = E\left[S_{R,I}\left|Y_{1,(R,I)},...,Y_{M,(R,I)}\right.\right] = \frac{\int_{-\infty}^{\infty} S_{R,I}\, p\left(Y_{1,(R,I)},...,Y_{M,(R,I)}\left|S_{R,I}\right.\right) p\left(S_{R,I}\right) dS_{R,I}}{\int_{-\infty}^{\infty} p\left(Y_{1,(R,I)},...,Y_{M,(R,I)}\left|S_{R,I}\right.\right) p\left(S_{R,I}\right) dS_{R,I}}, \tag{3}$$

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 spectral amplitude A and spectral phase α.

A Statistical Models

Based on the form of the distributions given in [2], Gaussian models are assumed for both the speech prior likelihood

$$p\left(S_{R,I}\right) = \frac{1}{\sqrt{\pi}\sigma_S} \exp\left(-\frac{S_{R,I}^2}{\sigma_S^2}\right) \tag{4}$$

and noise likelihood

$$p\left(Y_{i,(R,I)}\left|S_{R,I}\right.\right) = \frac{1}{\sqrt{\pi}\sigma_{N_i}} \exp\left(-\frac{\left(Y_{i,(R,I)} - c_i S_{R,I}\right)^2}{\sigma_{N_i}^2}\right), \tag{5}$$

where σ_S^2 and $\sigma_{N_i}^2$ are the speech and noise spectral variances. Since the MMSE estimator in (3) consists of a noise likelihood with M noisy microphone observations $\left\{Y_{1,(R,I)}, Y_{2,(R,I)},...,Y_{M,(R,I)}\right\}$ conditioned on the true real and imaginary spectral components $S_{R,I}$, the noise likelihood in (5) must account for all the available information, not simply at the i th microphone. Under the assumption of a diffuse noise field [21], the spectral real and imaginary noise components are uncorrelated as shown in

$$p\left(Y_{1,(R,I)},...,Y_{M,(R,I)}\left|S_{R,I}\right.\right) = \prod_{i=1}^{M} p\left(Y_{i,(R,I)}\left|S_{R,I}\right.\right) = \prod_{i=1}^{M} \frac{1}{\sqrt{\pi}\sigma_{N_i}} \exp\left(-\sum_{i=1}^{M} \frac{\left(Y_{i,(R,I)} - c_i S_{R,I}\right)^2}{\sigma_{N_i}^2}\right). \tag{6}$$

For the distributed microphone MMSE estimator that will be derived from (3), the relationship in (6) allows for the estimation of the noise statistics at each of the corresponding microphones.

B Optimal Multichannel MMSE Estimator

By substitution of Gaussian statistical models for the speech prior (4) and noise likelihood (6), the MMSE estimator in (3) is written as

$$\hat{S}_{R,I} = \frac{\int_{-\infty}^{\infty} S_{R,I} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I}\left(\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}\right)\right) dS_{R,I}}{\int_{-\infty}^{\infty} \exp\left(-S_{R,I}^2 \frac{1}{\lambda} + 2S_{R,I}\left(\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}\right)\right) dS_{R,I}} = \frac{P_{R,I}}{Q_{R,I}}, \tag{7}$$

where

$$\frac{1}{\lambda} = \frac{1}{\sigma_S^2} + \sum_{i=1}^{M} \frac{c_i^2}{\sigma_{N_i}^2}. \tag{8}$$

After splitting the integral in both the numerator and denominator in (7) each into two separate integrals and utilizing the relationship 3.462.1 in [22], the integration over $S_{R,I}$ produces the results

$$P_{R,I} = \left(2\frac{1}{\lambda}\right)^{-1} \exp\left(\frac{\left(\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}\right)^2}{2\frac{1}{\lambda}}\right)\left[D_{-2}\left(\frac{-\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right) - D_{-2}\left(\frac{\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)\right] \tag{9}$$

and

$$Q_{R,I} = \left(2\frac{1}{\lambda}\right)^{-\frac{1}{2}} \exp\left(\frac{\left(\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}\right)^2}{2\frac{1}{\lambda}}\right)\left[D_{-1}\left(\frac{\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right) + D_{-1}\left(\frac{-\sqrt{2}\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}}\right)\right], \tag{10}$$

where $D_{\bullet}(\bullet)$ is the parabolic cylinder function defined by 9.240 in [22] and

$$\left(2\frac{1}{\lambda}\right)^{-\frac{1}{2}} = \frac{\sqrt{2}}{2}\left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i}\right)^{\frac{1}{2}} \tag{11}$$

with $\sigma_{S_i}^2 = c_i^2 \sigma_S^2$. The arguments to the parabolic cylinder functions are simplified to and defined as

$$\pm\sqrt{2}\frac{\sum_{i=1}^{M} \frac{c_i Y_{i,(R,I)}}{\sigma_{N_i}^2}}{\left(\frac{1}{\lambda}\right)^{\frac{1}{2}}} = \pm\sqrt{2}\frac{\sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} Y_{i,(R,I)}}{\left(1+\sum_{i=1}^{M}\xi_i\right)^{\frac{1}{2}}} = N_{(R,I)\pm} \tag{12}$$

using the same notation as in [2]. Through the substitution of the constant term and definition in (11) and (12), the MMSE estimator in (7) is rewritten as

$$\hat{S}_{R,I} = \frac{\sqrt{2}}{2}\left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i}\right)^{\frac{1}{2}}\left[\frac{D_{-2}\left(N_{(R,I)-}\right) - D_{-2}\left(N_{(R,I)+}\right)}{D_{-1}\left(N_{(R,I)+}\right) + D_{-1}\left(N_{(R,I)-}\right)}\right] = \frac{\sqrt{2}}{2}\left(\frac{\sigma_S^2}{1+\sum_{i=1}^{M}\xi_i}\right)^{\frac{1}{2}} N_{(R,I)+}. \tag{13}$$

By simplification of (13), the closed-form multichannel Wiener filter solution $\hat{S}_{R,I}$ is represented as

$$\hat{S}_{R,I} = \frac{\sigma_S \sum_{i=1}^{M} \frac{\sqrt{\xi_i}}{\sigma_{N_i}} Y_{i,(R,I)}}{1+\sum_{i=1}^{M} \xi_i}, \tag{14}$$

which is applied to the real and imaginary spectral components of the noisy observation signals for distributed microphones. From the form of (14), it 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 $M = 1$, the MMSE estimate for S simplifies as

$$\hat{S} = \left|\hat{S}\right| e^{j\angle\hat{S}} = \hat{S}_R + j\hat{S}_I = \frac{\xi}{1+\xi} Y_R + j\frac{\xi}{1+\xi} Y_I = \left(\frac{\xi}{1+\xi}\left|Y\right|\right) e^{j\angle Y}, \tag{15}$$

which is the single channel noise reduction Wiener filter as stated in [2].

4. Simulation Experiments and Results

To evaluate the proposed optimal multichannel Wiener filter derived in (14), distributed multiple microphone noisy signals were simulated using the TIMIT [23] and NOISEX [24] corpora for constructing the attenuated clean speech signals and noise signals with input SNRs ranging from -10 dB to 10 dB. The analysis conditions consisted of frames of 256 samples (25.6 ms) with 50% overlap using Hanning windows. Noise estimation was performed on an initial silence of 5 frames. The decision-directed (DD) [3] smoothing approach was utilized to estimate ξ with $\alpha_{SNR} = 0.98$ using thresholds of $\xi_{\min} = 10^{-25/10}$ and $\gamma_{\min} = 40$. The microphones were assumed to be equal distance from the source signal, and the corresponding unity attenuation factors were estimated using the signal powers of the noisy signals across an entire utterance [25]. Objective measures of SSNR [26], LLR [27], and PESQ [28] were utilized to measure the noise reduction, speech distortion, and overall quality [10] averaged over ten enhanced signals, which were reconstructed using the overlap-add technique. Figure 1 shows the SSNR improvement, LLR output, and PESQ output as a function of the number of microphones in the array, where LLR (range of 0-2; lower scores indicate better performance) and PESQ (range of 0.5-4.5; higher scores indicate better performance).

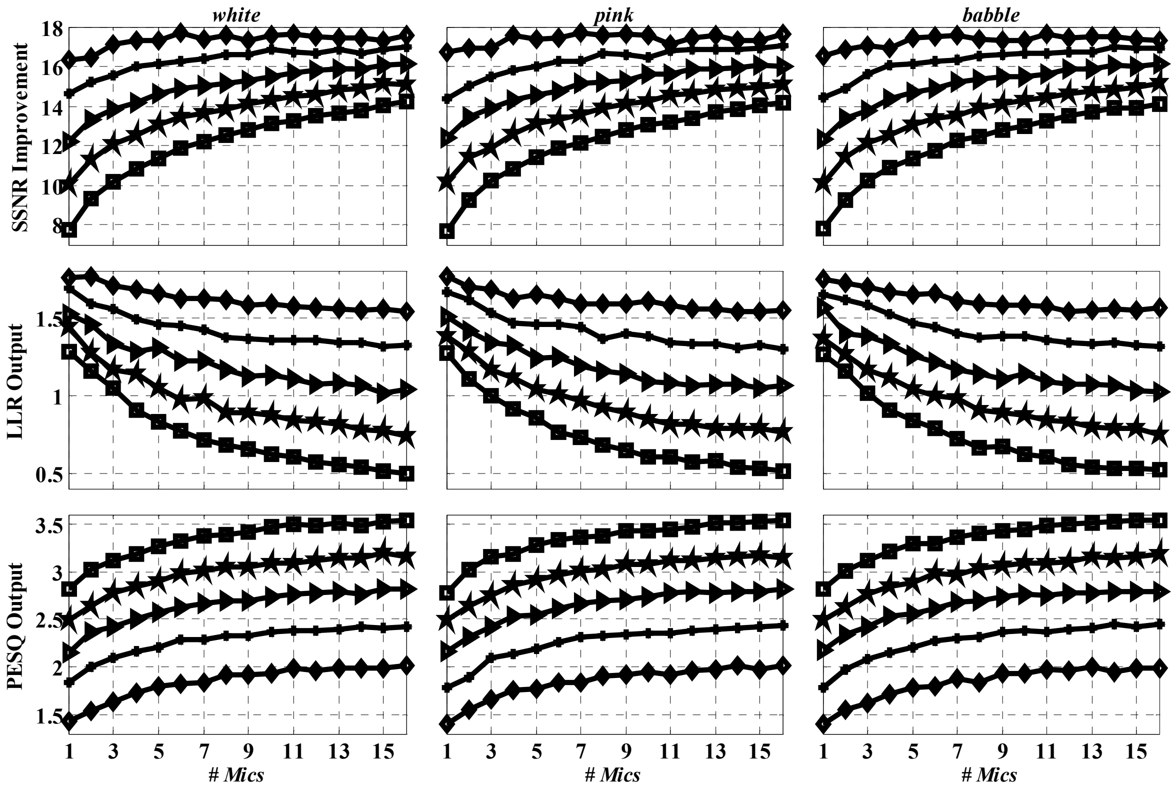

Figure 1. *SSNR Improvement, LLR Output, and PESQ Output for the Multichannel Wiener Filter (-10 dB = diamond, -5 dB = asterisk, 0 dB = right-pointing triangle, 5 dB = five-pointed star, and 10 dB = square).*

The multichannel Wiener filter achieved significant increases in noise reduction, decreases in speech distortion, and increases in overall quality at middle to higher input SNRs (0 dB – 10 dB) across the three different noise types for increasing number of microphones over the baseline single channel Wiener filter. In specific terms, the gains were approximately 4 dB, 5 dB, and 6 dB (SSNR improvement); 0.5, 0.7, and 0.8 (LLR output); and 0.8, 0.8, and 0.8 (PESQ output) for the 0 dB – 10 dB input SNR cases with the inclusion of the additional microphone information. By further examining the trends in the figure, it should be apparent that the performance of the optimal filter was robust to the various input SNR levels and different noise types. The multichannel Wiener filter simply functions well in a variety of noisy environments. At the lower input SNRs (-10 dB and -5 dB), the gains for the SSNR improvement (1 dB and 2 dB), LLR output (0.2 and 0.3), and PESQ output (0.5 and 0.7) were not as pronounced as with the middle to higher input SNRs (0 dB – 10 dB). The reasons are that the filter requires estimation of the noise and attenuation factors at each of the microphones, which are more difficult to estimate in the noisier of the two input SNR groups. Overall, the multichannel Wiener filter shows distinct performance benefits with the incorporation of the additional microphones measured by the SSNR, LLR, and PESQ objective measures.

5. Conclusion

In this letter, the multichannel Wiener filter was derived for multichannel speech enhancement in the distributed microphone paradigm. The emphasis of this work was to illustrate that the inclusion of the additional microphone information provides increases in noise reduction, decreases in speech distortion, and increases in overall speech quality as measured by SSNR as well as the LLR and PESQ objective metrics. Based on the experimental results, the multichannel Wiener filter achieved increases of approximately 5.0 dB (SSNR improvement), 0.7 (LLR output), and 0.8 (PESQ output) averaged across the 0 dB, 5 dB, and 10 dB input SNRs compared to the single channel Wiener filter baseline with less noticeable gains at lower input SNRs.

References

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

[2] 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.

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

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

[5] 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.

[6] I. Andrianakis and P. R. White, "Speech Spectral Amplitude Estimators using Optimally-Shaped Gamma and Chi Priors," Speech Communication, pp. 1-14, 2009.

[7] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum Mean-Square Error Estimation of Discrete Fourier Coefficients with Generalized Gamma Priors," IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, pp. 1741-1752, 2007.

[8] E. Plourde and B. Champagne, "Auditory-Based Spectral Amplitude Estimators for Speech Enhancement," IEEE Transactions on Audio, Speech and Language Processing, vol. 16, pp. 1614-1623, 2008.

[9] 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.

[10] Y. Hu and P. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement," IEEE Transactions on Audio, Speech, and Language Processing, vol. 16, pp. 229-238, 2008.

[11] 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.

[12] R. C. Hendriks, R. Heusdens, U. Kjerns, and J. Jensen, "On Optimal Multichannel Mean-Squared Error Estimators for Speech Enhancement," IEEE Signal Processing Letters, vol. 16, pp. 885-888, 2009.

[13] M. B. Trawicki, "Distributed Multichannel Processing for Signal Enhancement," in Electrical and Computer Engineering. Milwaukee: Marquette University, 2009, pp. 228.

[14] 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.

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

[16] S. Doclo and M. Moonen, "On the Output SNR of the Speech-Distortion Weighted Multichannel Wiener Filter," IEEE Signal Processing Letters, vol. 12, pp. 809-811, 2005.

[17] A. Spriet, M. Moonen, and J. Wouters, "Spatially Pre-Processed Speech Distortion Weighted Multi Channel Wiener Filtering for Noise Reduction," Signal Processing, vol. 84, pp. 2367-2387, 2004.

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

[19] I. A. McCowan, "Robust Speech Recognition using Microphone Arrays." Brisbane, Australia: Queensland University of Technology, 2001.

[20] C. H. Knapp and G. C. Carter, "The Generalized Correlation Method for Estimation of Time Delay," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-24. New York, NY, 1976, pp. 320-327.

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

[22] I. S. Gradshteyn and Z. M. Ryzhik, Table of Integrals, Series, and Products, 5th Edition ed. New York City, NY: Academic, 1994.

[23] J. Garofolo, L. Lamel, and W. Fisher, "TIMIT Acoustic-Phonetic Continuous Speech Corpus." Gaithersburg, MD: Linguistic Data Consortium, 1993.

[24] 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.

[25] M. B. Trawicki and M. T. Johnson, "Optimal Distributed Microphone Phase Estimation," presented at International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, R. O. C., 2009.

[26] P. E. Papamichalis, Practical Approaches to Speech Coding. New York, NY: Prentice-Hall, 1987.

[27] S. R. Quackenbush, I. T. P. Barnwell, and M. A. Clements, Objective Measures of Speech Quality. New York: Prentice-Hall, 1998.

[28] ITU-T, "Recommendation P.862: Perceptual Evaluation of Speech Quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," 2001.