

On Optimal Multichannel Mean-Squared Error Estimators for Speech Enhancement

Richard C. Hendriks, Richard Heusdens, Ulrik Kjems, and Jesper Jensen

Abstract—In this letter we present discrete Fourier transform (DFT) domain minimum mean-squared error (MMSE) estimators for multichannel noise reduction. The estimators are derived assuming that the clean speech magnitude DFT coefficients are generalized-Gamma distributed. We show that for Gaussian distributed noise DFT coefficients, the optimal filtering approach consists of a concatenation of a minimum variance distortionless response (MVDR) beamformer followed by well-known single-channel MMSE estimators. The multichannel Wiener filter follows as a special case of the presented MSE estimators and is in general suboptimal. For non-Gaussian distributed noise DFT coefficients the resulting spatial filter is in general nonlinear with respect to the noisy microphone signals and cannot be decomposed into an MVDR beamformer and a post-filter.

Index Terms—MMSE, multichannel, noise reduction.

I. INTRODUCTION

A COMMON way to make speech processing applications robust against environmental noise, i.e., increase speech quality, intelligibility and listening comfort, is to equip these applications with a noise reduction algorithm. Dependent on the specific application, noise reduction can be applied by exploiting single-channel techniques, e.g., [1]–[3], or multichannel techniques, e.g., [4], [5]. Multichannel methods allow for spatial filtering and therefore they tend to obtain superior quality over single-channel noise reduction methods.

Bayesian estimators [6], e.g., minimum mean-squared error (MMSE) estimators, have played an important role for single-channel methods, e.g., [1], [2], [7]–[9], as well as multichannel methods, e.g., [10]–[12]. A well-known multichannel estimator is the multichannel Wiener filter (MWF), e.g., [10], [13]. Often, the MWF is claimed to be mean-squared error (MSE) optimal, e.g., [11], [14], [15]. However, as the MWF can be derived using only second order moments of the signals involved, one would expect that the MWF is only truly MSE optimal for signals whose higher-order (beyond second) cumulants are zero, e.g., when both speech and noise are assumed to be Gaussian distributed. As histograms have demonstrated, the distribution of speech in time-domain and various transform

domains is not Gaussian, but rather super-Gaussian, e.g., [2], [8], [16], it seems plausible that the MWF is suboptimal, and performance would be improved by multichannel estimators derived under such non-Gaussian distributions.

In this letter we consider discrete Fourier transform (DFT) domain multichannel MSE estimators for speech enhancement under general distributional assumptions. We show that under certain realistic assumptions on the distribution of the noise DFT coefficients, that the MMSE estimator decomposes into a concatenation of a linear spatial filter and a, generally nonlinear, single-channel filter. More specifically, in this general case, the optimal filtering approach consists of a minimum variance distortionless response (MVDR) beamformer followed by a single-channel spectral enhancement scheme from a class of well-established MMSE estimators. The fact that the optimal filtering strategy involves two well-known estimators makes it straightforward to gauge the possible performance gain of the optimal strategy over the MWF.

Moreover, we show that when the noise DFT coefficients are modelled using more general distributional models, this decomposition does in general not hold anymore; the resulting spatial filter becomes nonlinear with respect to the noisy microphone signals.

II. NOTATION AND BASIC ASSUMPTIONS

We assume that each of the N noisy microphone signals is windowed and transformed to the DFT domain, leading to noisy DFT coefficients $X_n(k, i)$, where $n \in \{1, \dots, N\}$ is the microphone number, k the frequency-bin index and i the time-frame index. Let $S_n(k, i)$ and $V_n(k, i)$ denote clean speech and noise DFT coefficients, respectively. The DFT coefficients are assumed to be random variables, indicated by upper case letters. Their corresponding realizations are indicated by lower case letters. Furthermore, bold faced letters indicate the use of matrices. We assume that the speech and noise DFT coefficients are independent and additive i.e.,

$$X_n(k, i) = S_n(k, i) + V_n(k, i). \quad (1)$$

We assume the coefficients to be independent across time and frequency, which allows us to neglect time- and frequency-indices for ease of notation. For the complex speech DFT coefficients S_n , a polar representation will be used for mathematical convenience, i.e., $S_n = A_n e^{j\Phi_n}$, where A_n and Φ_n are the magnitude and phase of S_n , respectively. We assume that there is a single target speaker whose acoustic path to the N microphones is modelled by the frequency dependent propagation vector $\mathbf{d} \in \mathbb{C}^N$; consequently, the clean speech DFT coefficients $\mathbf{S} = [S_1, \dots, S_N]^T$ observed at each microphone are

Manuscript received May 06, 2009; revised June 17, 2009. First published June 26, 2009; current version published July 29, 2009. This work was supported by the Oticon foundation and the Dutch Technology Foundation STW. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrizio Campisi.

R. C. Hendriks and R. Heusdens are with the Department of Mediamatics, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: R.C.Hendriks@TUDelft.NL).

U. Kjems and J. Jensen are with Oticon A/S, 2765 Smørum, Denmark.
Digital Object Identifier 10.1109/LSP.2009.2026205

given by $S\mathbf{d}$, where $S = Ae^{j\Phi} \in \mathbb{C}$ is the clean speech DFT at the target speaker location. Let $\mathbf{X} \in \mathbb{C}^N$ be a vector containing the N noisy microphone DFT coefficients, i.e., $\mathbf{X} = [X_1, \dots, X_N]^T$. Similarly we define $\mathbf{V} \in \mathbb{C}^N$ as the vector consisting of the noise DFT coefficients at the N microphones, such that

$$\mathbf{X} = \mathbf{S} + \mathbf{V} = S\mathbf{d} + \mathbf{V}. \quad (2)$$

Further, let $\Sigma \in \mathbb{C}^{N \times N}$ be the noise correlation matrix defined as $\Sigma = E[\mathbf{V}\mathbf{V}^H]$, with E the statistical expectation operator.

III. OPTIMAL MULTICHANNEL MSE ESTIMATORS

The estimator of $h(S)$, say $\widehat{h(S)}$, that minimizes the MSE is given by $\widehat{h(S)} = E[h(S)|\mathbf{X}]$ [6], where $h(\cdot)$ is introduced to allow for MSE estimation of functions $h(\cdot)$ of the random variable S , like $h(S) = |S|$. Using Bayes rule, and the fact that the speech phase Φ is uniformly distributed and independent of the speech magnitude A [9], we obtain

$$E[h(S)|\mathbf{x}] = \frac{\int_A \int_{\Phi} h(ae^{j\phi}) f_{\mathbf{X}|A,\Phi}(\mathbf{x}|a, \phi) f_A(a) d\phi da}{\int_A \int_{\Phi} f_{\mathbf{X}|A,\Phi}(\mathbf{x}|a, \phi) f_A(a) d\phi da}. \quad (3)$$

A. Assuming Complex-Gaussian Noise DFT Coefficients

The distribution $f_{\mathbf{X}|A,\Phi}$ is determined by the distribution of the vector \mathbf{V} of noise DFT coefficients. For noise sources with a relatively short time-span of dependency [17], it is realistic to model $f_{\mathbf{X}|A,\Phi}$ as a multivariate complex-Gaussian distribution, i.e.,

$$f_{\mathbf{X}|A,\Phi}(\mathbf{x}|a, \phi) = \frac{1}{\pi^N |\Sigma|} \exp[-(\mathbf{x} - S\mathbf{d})^H \Sigma^{-1} (\mathbf{x} - S\mathbf{d})] \quad (4)$$

with $|\Sigma|$ the determinant of Σ . This assumption was experimentally verified by Lotter in [8] for the one-dimensional case of $f_{\mathbf{X}|A,\Phi}$ and easily extends to the multivariate case.

1) *Complex-DFT Multichannel MMSE Estimator:* Substitution of (4) into (3), setting $h(\cdot)$ to $h(S) = S$ and using [18, Eq. 3.937.2] to compute the integral over ϕ , we obtain

$$E[S|\mathbf{x}] = e^{j\angle z(\mathbf{x})} t(z(\mathbf{x})) \quad (5)$$

with

$$z(\mathbf{x}) = \frac{\mathbf{d}^H \Sigma^{-1} \mathbf{x}}{\mathbf{d}^H \Sigma^{-1} \mathbf{d}} \quad (6)$$

and where $\angle z(\mathbf{x})$ denotes the phase of $z(\mathbf{x})$, and

$$t(z(\mathbf{x})) = \frac{\int_A a f_A(a) e^{-a^2 \mathbf{d}^H \Sigma^{-1} \mathbf{d}} \mathcal{I}_1(2a \mathbf{d}^H \Sigma^{-1} \mathbf{d} |z(\mathbf{x})|) da}{\int_A f_A(a) e^{-a^2 \mathbf{d}^H \Sigma^{-1} \mathbf{d}} \mathcal{I}_0(2a \mathbf{d}^H \Sigma^{-1} \mathbf{d} |z(\mathbf{x})|) da} \quad (7)$$

where $\mathcal{I}_n(\cdot)$ is the modified Bessel function of the first kind and order n . We recognize (6) as the MVDR beamformer applied on the noisy observation \mathbf{x} . From (5)–(7) we can draw several conclusions. First, (5) is a function of \mathbf{x} , only through the function $z(\mathbf{x})$. This is due to the assumption that the noise DFT

vector \mathbf{V} has a multivariate complex-Gaussian distribution. Secondly, from (5)–(7) we identify that the optimal estimator consists of a concatenation of two processing steps; an MVDR beamformer applied to \mathbf{x} (6), followed by post-processing of the MVDR beamformer output $z(\mathbf{x})$ in (7). Thirdly, observing that $\mathbf{d}^H \Sigma^{-1} \mathbf{d}$ is real, it follows that the function $t(z(\mathbf{x}))$ is real. Therefore, we can conclude from (5) that the phase of the MMSE estimator equals the phase of the MVDR beamformer output and its magnitude is determined by $t(z(\mathbf{x}))$.

To specify the post-processing applied in (7) to $z(\mathbf{x})$, we must specify $f_A(a)$. As in [9] we assume that the magnitude A is generalized-Gamma distributed, i.e.,

$$f_A(a) = \frac{\gamma \beta^\nu}{\Gamma(\nu)} a^{\nu-1} \exp(-\beta a^\gamma), \quad \beta > 0, \nu > 0, a \geq 0 \quad (8)$$

where $\Gamma(\cdot)$ is the Gamma function and where we set $\gamma = 2$ from which it follows [9] that $\beta = \nu/\sigma_S^2$, with $\sigma_S^2 = E[|S|^2]$. Let \mathcal{M} denote the confluent hypergeometric function [18]. Substitution of (8) into (5) and subsequently using [18, Eqs. 6.643.2, 9.220.2] then leads to

$$E[S|\mathbf{x}] = \frac{\nu \sigma_S^2}{\nu(\mathbf{d}^H \Sigma^{-1} \mathbf{d})^{-1} + \sigma_S^2} \times \frac{\mathcal{M}(\nu + 1, 2, P[z(\mathbf{x})])}{\mathcal{M}(\nu, 1, P[z(\mathbf{x})])} z(\mathbf{x}) \quad (9)$$

with

$$P[z(\mathbf{x})] = \frac{\sigma_S^2 \mathbf{d}^H \Sigma^{-1} \mathbf{d} |z(\mathbf{x})|^2}{\nu(\mathbf{d}^H \Sigma^{-1} \mathbf{d})^{-1} + \sigma_S^2}. \quad (10)$$

Equation (9) can be recognized as the multichannel extension of the single-channel MMSE estimator presented in [19]. The noisy input in (9) is the MVDR beamformer output $z(\mathbf{x})$. The noise in the MVDR beamformer output remains Gaussian, since \mathbf{V} is assumed to be complex-Gaussian distributed and the MVDR beamformer is linear and deterministic. From (9) we can conclude that the MWF is suboptimal in general. An exception is the case in which $\nu = 1$, i.e., the complex speech DFT coefficients are assumed to have a complex-Gaussian distribution, for which (9) reduces to the MWF solution. Since it is well known that the distribution of speech DFT coefficients tends to be super-Gaussian, see e.g., [2], [8], other choices for f_A that have a better match with actual speech data will lead to a solution that is closer to the optimal estimator.

Experimental Validation: To experimentally validate the suboptimality of the MWF we conduct an experiment where the performance of the multichannel MMSE estimator in (9) is compared to the MWF and the single-channel MMSE estimator presented in [19] that is derived under exactly the same distributional assumptions as the estimator in (9). The comparison is performed in terms of the quantity that the MMSE estimators optimize for, i.e., mean-squared error (MSE), estimated as

$$\text{MSE} = \frac{1}{KI} \sum_{\forall(k,i)} |s(k,i) - E[S(k,i)|\mathbf{x}(k,i)]|^2 \quad (11)$$

where K and I denote the total number of frequency bins and time-frames, respectively. We consider a dual-microphone end-fire array setup with microphone distance of 1 cm. The target

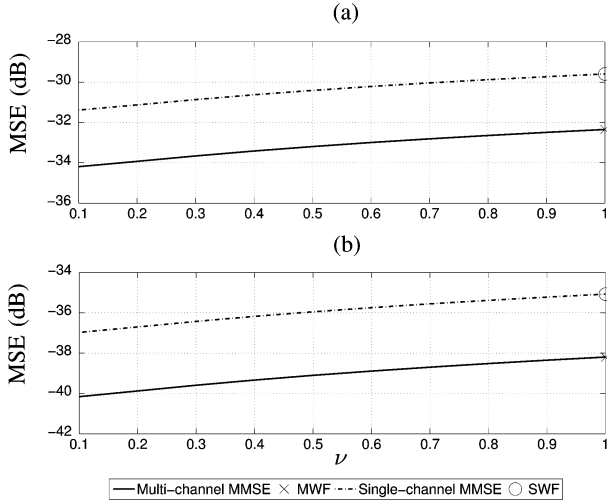


Fig. 1. Performance in terms of MSE (dB) for speech degraded by two white-noise point sources positioned at 40 and 140 degrees at an SNR of (a) 5 dB (b) 15 dB.

speech source is positioned at zero degrees and degraded by two white-noise point sources positioned at 40 and 140 degrees. The noise correlation matrix Σ for the multichannel MMSE estimator in (9) and the noise PSD for the single-channel MMSE estimator in [19] are estimated using an ideal voice activity detector and the propagation vector \mathbf{d} is assumed to be given. The estimators are applied in the Fourier domain on a frame-by-frame basis to time-frames of 32 ms taken with 50% overlap. Evaluations are performed using a data-base consisting Danish speech spoken by nine female and eight male speakers and sampled with a sampling frequency of 8 kHz. In Fig. 1(a) and (b) the performance in terms of MSE is given for SNRs of 5 dB and 15 dB, respectively, as a function of the ν -parameter. The MWF that appears as a special case of the in (9) presented estimator for $\nu = 1$, is indicated in Fig. 1 by the symbol \times . The single-channel Wiener filter (SWF), a special case of the single-channel MMSE estimators presented in [19], is indicated by the symbol \circ . Clearly, we see that that the MWF is suboptimal and leads to a MSE that is in the order of 2 dB larger compared to multichannel MMSE estimators derived under super-Gaussian distributions with ν in the order of 0.1. The improvement of the multichannel MMSE estimators over the single-channel MMSE estimators under exactly the same distributional assumptions, i.e., the same ν -parameter, is in this situation in the order of 2.5–3 dB, but will generally depend on the number, position and characteristics of the noise sources and the geometry of the microphone setup.

2) *Magnitude-DFT Multichannel MMSE Estimator:* Besides complex DFT estimators, magnitude-DFT estimators also received a lot of attention for noise reduction, e.g., [1], [8]. Choosing $h(S) = |S|$ in (3) and applying derivations along the same line as is done to obtain (5), we obtain the multichannel magnitude-DFT estimator $E[A|\mathbf{X}]$, that is

$$E[A|\mathbf{x}] = \frac{\Gamma(\nu + \frac{1}{2})}{\Gamma(\nu)} \sqrt{\frac{\sigma_S^2}{\nu + \sigma_S^2(\mathbf{d}^H \Sigma^{-1} \mathbf{d})}}$$

$$\times \frac{\mathcal{M}(\nu + 1/2, 1, P[z(\mathbf{x})])}{\mathcal{M}(\nu, 1, P[z(\mathbf{x})])} \quad (12)$$

with $P[z(\mathbf{x})]$ as defined in (10). Equation (12) is the multi-channel extension of the single-channel MMSE estimator presented in [9]. In analogy with (9), the MMSE magnitude-DFT estimator can be decomposed into an MVDR beamformer represented by $z(\mathbf{x})$ and a post-filter that is well known from single-channel MMSE estimation, e.g., [9]. The fact that the optimal filtering approach consists of an MVDR beamformer followed by a (potentially nonlinear) single-channel post-processor was also observed by Balan [20], who derived explicit expressions for the case of complex-Gaussian noise DFT coefficients and complex-Gaussian speech DFT coefficients for $h(S) = |S|$, $h(S) = \log |S|$ and $h(S) = S/|S|$.

B. Assuming Non-Gaussian Noise DFT Coefficients

The fact that the multichannel MMSE estimators can be decomposed into a concatenation of an MVDR beamformer and a single-channel post-processor holds as long as \mathbf{V} is assumed to be multivariate complex-Gaussian distributed, but does not hold in general. More specifically, we can generalize the results in Section III-A by assuming that \mathbf{V} is distributed according to a multivariate complex-Gaussian mixture, i.e.,

$$f_{\mathbf{X}|A, \Phi}(\mathbf{x}|a, \phi) = \sum_{m=1}^M \frac{c_m}{\pi^N |\Sigma_m|} e^{[-(\mathbf{x} - \mathbf{s}\mathbf{d})^H \Sigma_m^{-1} (\mathbf{x} - \mathbf{s}\mathbf{d})]} \quad (13)$$

with M the number of components and c_m positive weighting factors that satisfy $\sum_{m=1}^M c_m = 1$. Furthermore, Σ_m are the variances of the individual components and satisfy $\sum_{m=1}^M \Sigma_m = \Sigma$. Such a distribution is useful to model noise DFT coefficients from sources with a rather long time-span of dependency, e.g., babble- and fan-noise, for which histograms of DFT coefficients deviate somewhat from Gaussianity [8]. Under these more general distributional assumptions, substitution of (8) and (13) into (3), setting $h(S) = S$ and using [18, Eqs. 3.937.2, 6.643.2, and 9.220.2] leads to

$$E[S|\mathbf{x}] = \nu \frac{\sum_{m=1}^M \frac{c_m Q_m}{|\Sigma_m|} e^{[-\mathbf{x}^H \Sigma_m^{-1} \mathbf{x}]} \frac{\sigma_S^2 z_m(\mathbf{x}) \mathcal{M}(\nu+1, 2, P_m)}{\nu (\mathbf{d}^H \Sigma_m^{-1} \mathbf{d})^{-1} + \sigma_S^2}}{\sum_{m=1}^M \frac{c_m Q_m}{|\Sigma_m|} e^{[-\mathbf{x}^H \Sigma_m^{-1} \mathbf{x}]} \mathcal{M}(\nu, 1, P_m)} \quad (14)$$

with

$$z_m(\mathbf{x}) = \frac{\mathbf{d}^H \Sigma_m^{-1} \mathbf{x}}{\mathbf{d}^H \Sigma_m^{-1} \mathbf{d}}, \quad Q_m = (\nu + \mathbf{d}^H \Sigma_m^{-1} \mathbf{d} \sigma_S^2)^{-\nu}$$

and

$$P_m = \frac{\sigma_S^2 \mathbf{d}^H \Sigma_m^{-1} \mathbf{d} |z_m(\mathbf{x})|^2}{\nu (\mathbf{d}^H \Sigma_m^{-1} \mathbf{d})^{-1} + \sigma_S^2}. \quad (15)$$

We see that under this non-Gaussian noise DFT assumption, the multichannel MMSE estimator can *not* be decomposed into an MVDR beamformer and a single-channel post-processor; the spatial filter becomes nonlinear with respect to \mathbf{x} . In a similar way, the magnitude estimator $E[A|\mathbf{x}]$ based on (8) and (13) is obtained by setting $h(S)$ in (3) to $h(S) = |S|$. This magnitude

estimator can also not be decomposed into an MVDR beamformer and a single-channel magnitude-DFT MMSE estimator as a post-processor.

IV. CONCLUSION

In this letter we treated the MSE optimal multichannel filtering problem. Specifically, we considered DFT-domain multichannel estimators for speech enhancement. The estimators are derived assuming that the speech magnitude-DFT coefficients are generalized-Gamma distributed; this assumption has been justified in several studies, e.g., [8], [9]. When the noise DFT coefficients are complex-Gaussian distributed, we can conclude that, independent of the distribution of the speech magnitude-DFT coefficients, the optimal filtering strategy is a concatenation of two well-known methods; a linear spatial filter (MVDR) and a, generally nonlinear, single-channel filter applied to the MVDR beamformer output, i.e., an intuitively pleasing and straightforward concatenation of very well-known techniques. Further, we can conclude that

- the MWF is generally suboptimal;
- the performance loss in using the MWF instead of the optimal filter is determined by the performance differences of the single-channel post-processors; for speech signals this gap is around 2 dB in terms of MSE [19];
- the MMSE estimator can be derived using the distribution $f_A(a)$ observed at the target source location, independent of the acoustical situation (given that the assumed signal model is valid). As a consequence, speech targets can be modelled appropriately using super-Gaussian densities.

Furthermore, when noise DFT coefficients cannot be assumed to be complex-Gaussian distributed, we can conclude that in general

- the optimal spatial filter is nonlinear;
- it is not possible to decompose the estimator into an MVDR beamformer and a single-channel post-processor.

REFERENCES

- [1] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.
- [2] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 845–856, Sep. 2005.
- [3] *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Berlin, Germany: Springer, 2005.
- [4] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, Oct. 1987.
- [5] *Microphone Arrays*, M. Brandstein and D. Ward, Eds. New York: Springer, 2001.
- [6] H. L. van Trees, *Detection, Estimation and Modulation Theory*. New York: Wiley, 1968, vol. 1.
- [7] Y. Ephraim, "Statistical-model-based speech enhancement systems," *Proc. IEEE*, vol. 80, no. 10, pp. 1526–1555, Oct. 1992.
- [8] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 7, pp. 1110–1126, 2005.
- [9] 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 Trans. Audio Speech Lang. Process.*, vol. 15, no. 6, pp. 1741–1752, Aug. 2007.
- [10] L. W. Brooks and I. S. Reed, "Equivalence of the likelihood ratio processor, the maximum signal-to-noise ratio filter and the Wiener filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-8, no. 5, pp. 690–692, Sept. 1972.
- [11] S. Doclo, "Multi-Microphone Noise Reduction and Dereverberation Techniques for Speech Applications," Ph.D. thesis, Katholieke Univ. Leuven, Leuven, Belgium, 2003.
- [12] S. Srinivasan, R. Aichner, W. B. Kleijn, and W. Kellermann, "Multi-channel parametric speech enhancement," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 304–307, May 2006.
- [13] L. J. Griffiths, "A comparison of multidimensional Wiener and maximum-likelihood filters for antenna arrays," *Proc. IEEE*, vol. 55, no. 11, pp. 2045–2047, Nov. 1967.
- [14] J. C. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 14, no. 4, pp. 1218–1233, July 2006.
- [15] S. Leukimmiatis, D. Dimitriadis, and P. Maragos, "An optimum microphone array post-filter for speech applications," in *Interspeech*, Sep. 2006, pp. 2142–2145.
- [16] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 204–207, Jul. 2003.
- [17] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia, PA: SIAM, 2001.
- [18] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York: Academic, 2000.
- [19] R. C. Hendriks, J. S. Erkelens, and R. Heusdens, "Comparison of complex-DFT estimators with and without the independence assumption of real and imaginary parts," in *IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2008, pp. 4033–4036.
- [20] R. Balan and J. Rosca, "Microphone array speech enhancement by Bayesian estimation of spectral amplitude and phase," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2002.