ON ROBUSTNESS OF MULTI-CHANNEL MINIMUM MEAN-SQUARED ERROR ESTIMATORS UNDER SUPER-GAUSSIAN PRIORS

Richard C. Hendriks and Richard Heusdens

Delft University of Technology
2628 CD Delft, The Netherlands
R.C.Hendriks@tudelft.nl, R.Heusdens@tudelft.nl

Jesper Jensen

Oticon A/S
2765 Smørum, Denmark
jsj@oticon.dk

ABSTRACT

The use of microphone arrays in speech enhancement applications offer additional features, like directivity, over the classical single-channel speech enhancement algorithms. An often used strategy for multi-microphone noise reduction is to apply the multi-channel Wiener filter, which is often claimed to be mean-squared error optimal. However, this is only true if the estimator is constrained to be linear, or, if the speech and noise process are assumed to be Gaussian. Based on histograms of speech DFT coefficients it can be argued that optimal multi-channel minimum mean-squared error (MMSE) estimators should be derived under super-Gaussian speech priors instead. In this paper we investigate the robustness of these estimators when the steering vector is affected by estimation errors. Further, we discuss the sensitivity of the estimators when the true underlying distribution of speech DFT coefficients deviates from the assumed distribution.

***Index Terms*—** speech enhancement, noise reduction, multi-channel, MMSE, super-Gaussian

1. INTRODUCTION

The trend that speech processing devices like mobile phones and hearing aids should work at anytime and anywhere increases the need for making these systems robust and to improve quality, listening comfort and intelligibility. This can be done by exploiting multi-microphone noise reduction techniques. An often used strategy is to process the noisy signal on a frame-by-frame basis in the spectral domain, e.g., by applying a discrete Fourier transform (DFT), and use the multi-channel Wiener filter (MWF), see e.g., [1][2], to estimate the clean speech DFT coefficients. The MWF is often claimed to be mean-squared error optimal [3][4][5]. However, this is only true if the estimator is constrained to be linear, or, if the speech and noise process are assumed to be Gaussian. However, studies on the distribution of speech in time-domain as well as in several transform domains [6][7][8] show that the observed distribution is not Gaussian but super-Gaussian instead. This indicates that for speech the MWF is not optimal, and that better minimum MSE (MMSE) estimators could be derived by exploiting super-Gaussian distributions as a model for the speech process.

The multi-channel MMSE estimators presented in [9] are based on the observation that the speech DFT coefficients follow a super-Gaussian distribution. Similar to the MWF, these estimators are functions of the steering or propagation vector [9]. In practice the steering vector is not known with certainty and often estimated. Although the estimators in [9] are derived under distributional assumptions that are in line with measured histograms of speech DFT coefficients and can as such be considered to be MSE optimal, it is yet unclear how the performance of these multi-channel estimators is affected when there is a mismatch between the (unknown) true steering vector and the assumed steering vector. Further, it is unknown how the performance of these estimators is affected when the true underlying distribution deviates somewhat from the assumed super-Gaussian distribution. For example, one would expect the distribution of speech DFT coefficients observed at the microphone to change due to propagation and room acoustics if the target speech source is located some distance away. In this paper we will therefore discuss the performance impact when there is a mismatch between the true and estimated steering vector, and when there is a mismatch between the true and assumed underlying speech DFT distribution.

The remaining sections of this paper are organized as follows. In Section 2 we introduce the notation and basic assumptions that are used in this paper. In Section 3 we give a brief review of multi-channel MMSE estimators under the assumptions that speech DFT coefficients are super-Gaussian distributed. Subsequently, in Section 4 we discuss robustness of these MMSE estimators with respect to the estimated steering vector. Then, in Section 5 sensitivity of multi-channel MMSE estimators with respect to the assumed prior distribution is discussed. Finally, in Section 6 conclusions are drawn.

2. NOTATION AND BASIC ASSUMPTIONS

We assume the availability of N noisy microphone signals that are windowed and transformed to the DFT domain on a frame-by-frame basis. The noisy DFT coefficients are denoted by $X_n(k,i)$, with $n \in \{1, ..., N\}$ the microphone index, k the frequency-bin index and i the time-frame index. 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 an additive noise model, i.e.,

$$X_n(k,i) = S_n(k,i) + V_n(k,i), \tag{1}$$

where $S_n(k,i)$ and $V_n(k,i)$ denote the clean speech and noise DFT coefficients, respectively, that are assumed to have zero-mean and are statistically independent from each other. Further, we assume the DFT coefficients to be independent across time and frequency and 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.,

The research is supported by the Oticon foundation and the Dutch Technology Foundation STW.

$S_n = A_n e^{j\Phi_n}$, 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 steering vector $\mathbf{d} \in \mathbb{C}^N$; consequently, the clean speech DFT coefficients $\mathbf{S} = [S_1, ..., S_N]^T$ observed at the microphone array are given by $S\mathbf{d}$, where $S = Ae^{j\Phi} \in \mathbb{C}$ is the clean speech DFT coefficient of the target speaker. Let $\mathbf{X} \in \mathbb{C}^N$ be a vector of N noisy microphone DFT coefficients, i.e., $\mathbf{X} = [X_1, ..., X_N]^T$. Similarly we define $\mathbf{V} \in \mathbb{C}^N$ as the vector consisting of the noise DFT coefficients at the N microphones, such that

$$\mathbf{X} = \mathbf{S} + \mathbf{V} = S\mathbf{d} + \mathbf{V}. \tag{2}$$

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

3. MULTI-CHANNEL MMSE ESTIMATOR UNDER SUPER-GAUSSIAN PRIOR

To facilitate the discussion in this paper, we first give a brief review of the distributional assumptions, derivation and main conclusions of the multi-channel MMSE estimators as given in [9].

3.1. Distributional Assumptions

From the measured histograms of speech DFT coefficients [10] it follows that the speech DFT-phase Φ is uniformly distributed and independent from the DFT-magnitude A, while real and imaginary parts of speech DFT coefficients are clearly dependent. Further, from speech DFT histograms published in [7][8] it follows that the distribution of speech DFT coefficients tends to be super-Gaussian, i.e., heavy tailed and more peaked than the Gaussian distribution.

Based on the above considerations we assume in our derivations that the phase Φ of speech DFT coefficients is uniformly distributed and independent from the speech DFT-magnitude A. To be in line with the fact that speech DFT coefficients are super-Gaussian distributed we model the speech DFT-magnitudes with a fairly general probability density function, i.e., the generalized Gamma distribution, which can model for certain parameter settings super-Gaussian data. The use of the generalized-Gamma distribution to model speech DFT-magnitudes has been justified in several studies, e.g., [8][10]. The generalized-Gamma distribution is given by

$$f_A(a) = \frac{\gamma\beta^\nu}{\Gamma(\nu)} a^{\gamma\nu-1} \exp(-\beta a^\gamma),\ \beta > 0,\ \nu > 0,\ a \geq 0, \tag{3}$$

where we set $\gamma = 2$ from which it follows that $\beta = \nu/\sigma_S^2$, with $\sigma_S^2 = E\left[|S|^2\right]$ [10]. Whether a distribution is super-Gaussian can be deduced from the kurtosis of the distribution. For a complex random variable S, kurtosis can be defined as [11]

$$\text{kurt}(S) = E\left[|S|^4\right] - 2E\left[|S|^2\right]^2 - |E\left[S^2\right]|. \tag{4}$$

From the distributional assumptions above it follows that when ν in Eq. (3) is chosen $0 < \nu < 1$, the kurtosis of the speech DFT coefficients S is positive for which the distribution of S is said to be super-Gaussian. For $\nu = 1$ the kurtosis is zero and the distribution of S is complex Gaussian.

The noise DFT coefficients are assumed to have a complex zero-mean Gaussian distribution. This is based on the fact that the time-span of dependency [12] is relatively low for many noise sources, see e.g., [8], where histograms of several noise sources were measured. Based on this distributional assumption, together with the assumed independence of speech and noise DFT coefficients, it follows that the distribution $f_{\mathbf{X}|A,\Phi}$ is multivariate Gaussian, i.e.,

$$f_{\mathbf{X}|A,\Phi}(\mathbf{x}|a,\phi) = \frac{1}{\pi^N|\mathbf{\Sigma}|} \exp\left[-\left(\mathbf{x} - s\mathbf{d}\right)^H \mathbf{\Sigma}^{-1}\left(\mathbf{x} - s\mathbf{d}\right)\right], \tag{5}$$

with $|\mathbf{\Sigma}|$ the determinant of $\mathbf{\Sigma}$.

3.2. MMSE Estimator

The estimator $\hat{S}$ that is MMSE optimal can be obtained by computing $\hat{S} = E\left[S|\mathbf{X}\right]$ [13]. Using Bayes' rule, and the assumptions that the speech DFT phase Φ is uniformly distributed and independent from the magnitude A, we obtain

$$E\left[S|\mathbf{x}\right] = \frac{\int_A \int_\Phi a e^{j\phi} f_{\mathbf{X}|A,\Phi}(\mathbf{x}|a,\phi) f_A(a) d\phi da}{\int_A \int_\Phi f_{\mathbf{X}|A,\Phi}(\mathbf{x}|a,\phi) f_A(a) d\phi da}. \tag{6}$$

Let $\mathcal{M}$ denote the confluent hypergeometric function [14]. Substitution of Eqs. (3) and (5) into Eq. (6) and using [14, 3.937.2,6.643.2 and 9.220.2] leads to

$$E\left[S|\mathbf{x}\right] = \frac{\nu\sigma_S^2}{\nu(\mathbf{d}^H\mathbf{\Sigma}^{-1}\mathbf{d})^{-1} + \sigma_S^2} \frac{\mathcal{M}(\nu+1, 2, P)}{\mathcal{M}(\nu, 1, P)} z(\mathbf{x}), \tag{7}$$

with

$$z(\mathbf{x}) = \frac{\mathbf{d}^H\mathbf{\Sigma}^{-1}\mathbf{x}}{\mathbf{d}^H\mathbf{\Sigma}^{-1}\mathbf{d}}, \tag{8}$$

and

$$P = \frac{\sigma_S^2 \mathbf{d}^H\mathbf{\Sigma}^{-1}\mathbf{d}|z(\mathbf{x})|^2}{\nu(\mathbf{d}^H\mathbf{\Sigma}^{-1}\mathbf{d})^{-1} + \sigma_S^2}. \tag{9}$$

Eq. (7) can be recognized as the multi-channel extension of the single-channel MMSE estimator presented in [15]. For detailed conclusions on this multi-channel MMSE estimator we refer to [9]. Here we only highlight three of these conclusions. First, since $z(\mathbf{x})$ is a sufficient statistic for $\mathbf{x}$, we have that $E\left[S|z(\mathbf{x})\right] = E\left[S|\mathbf{x}\right]$. Secondly, from Eqs. (7)-(9) we identity that the optimal estimator consists of a concatenation of two processing steps: a minimum variance distortionless response (MVDR) beamformer applied to $\mathbf{x}$, as given by Eq. (8), followed by post-processing of the MVDR-beamformer output $z(\mathbf{x})$, given by Eq. (7). The post-processor is generally non-linear and can be recognized as a single-channel complex-DFT estimator, see [15]. Third, we can conclude that the MWF is sub-optimal in general. The MWF follows as a special case when setting $\nu = 1$, i.e., the complex speech DFT coefficients are assumed to have a Gaussian distribution. As argued above, the distribution of speech DFT coefficients tends to be super-Gaussian, see e.g. [7][8], and other choices than $\nu = 1$ in Eq. (3) that have a better match with actual speech data will lead to a solution that is closer to the optimal estimator.

Notice that under other distributions for the noise DFT vector $\mathbf{V}$, e.g. when $\mathbf{V}$ is assumed to follow a multivariate Gaussian mixture, the MMSE estimator will in general be a function of $\mathbf{X}$ and cannot be decomposed into an MVDR beamformer and a single-channel post-processor [9].

4. ROBUSTNESS WITH RESPECT TO THE ESTIMATED STEERING VECTOR

From the estimator given in Eqs. (7)-(9) it follows that the optimal multi-channel MSE estimator consists of a single-channel MMSE post-processing in Eq. (7) of the MVDR-beamformer output $z(\mathbf{x})$ from Eq. (8). The spatial filter, i.e., MVDR beamformer in Eq. (8), is conditioned on exact knowledge of the steering vector $\mathbf{d}$ and has by definition a distortionless response in the direction of the target speaker. However, in practice $\mathbf{d}$ is unknown and has to be estimated instead.

A mismatch between $\mathbf{d}$ and the estimated steering vector, say $\mathbf{m}$, implies that the MVDR beamformer is steered in the wrong direction. Consequently, the filter z will not be the optimal (linear) spatial filter of $\mathbf{X}$ and is not distortionless with respect to S, i.e., $z(\mathbf{S}) \neq S$. What can then be said about optimality of the post-processor as given by Eq. (7)?

The post-processor in Eq. (7) is derived under certain distributional assumptions on the speech DFT coefficients S and the vector of noise DFT coefficients $\mathbf{V}$. Given $z(\mathbf{X})$, the form of the post-processor remains optimal as long as the distributions of both $z(\mathbf{V})$ and $z(\mathbf{S})$ are not affected (up to a possible real scaling) by the mismatch between $\mathbf{m}$ and $\mathbf{d}$. Since the noise DFT coefficients $\mathbf{V}$ are assumed to be zero-mean Gaussian distributed and the MVDR beamformer is linear and deterministic, the resulting random variable $z(\mathbf{V})$ remains zero-mean Gaussian. With respect to the distribution of the filtered speech coefficients $z(\mathbf{S})$, it holds by using Eq. (8), under the assumption that speech is stationary and time-frames are sufficiently long, that

$$z(\mathbf{S}) = \frac{\mathbf{m}^H \mathbf{\Sigma}^{-1} S\mathbf{d}}{\mathbf{m}^H \Sigma^{-1}\mathbf{m}} = S\alpha, \tag{10}$$

with $\alpha \in \mathbb{C}$, where $\alpha = 1$ if and only if $\mathbf{d} = \mathbf{m}$. Since the distribution of S is rotational invariant [8][10], the argument of α will not affect the distribution of $z(\mathbf{S})$. Hence, $z(\mathbf{S})$ has the same distribution as S, up to a possible scaling $|\alpha|$.

In conclusion, we can state that estimation errors on $\mathbf{d}$ have no effect on the shape of the distribution of $z(\mathbf{S})$ and $z(\mathbf{V})$. From this it can be deduced that the post-processor in Eq. (7) remains the optimal processor for the MVDR-beamformer output $z(\mathbf{X})$, regardless the mismatch between $\mathbf{d}$ and $\mathbf{m}$. However, it should be noticed that the mismatch between $\mathbf{d}$ and $\mathbf{m}$ in general changes the scale of the distributions of $z(\mathbf{S})$ and $z(\mathbf{V})$, and therefore it can change the signal-to-noise ratio (SNR) in $z(\mathbf{X})$.

5. ROBUSTNESS WITH RESPECT TO THE ASSUMED PRIOR

The estimator in Eqs. (7)-(9) is derived under a certain prior distribution for the speech DFT coefficients, which is specified by the ν-parameter in the distribution f_A in Eq. (3). In this section we investigate how sensitive the performance of the estimator is when the assumed distribution deviates from the true, but unknown, distribution. We do this by computing the MSE, that is

$$E_{\mathbf{X},S}\left[|S - E_{\mathrm{a}}\left[S|\mathbf{X}\right]|^2\right], \tag{11}$$

where $E_{\mathbf{X},S}$ is the expected value with respect to the random variables $\mathbf{X}$ and S and where the subscript a in $E_{\mathrm{a}}\left[S|\mathbf{X}\right]$ indicates that this is the conditional expectation under the assumed distribution $f_{A_{\mathrm{a}}}(a)$. In a similar way we use the subscript t to indicate the true distribution $f_{A_{\mathrm{t}}}(a)$, that is, $E_{\mathrm{t}}\left[S|z(\mathbf{X})\right]$, ν_{a} and ν_{t} indicate the conditional expectation under the true distribution $f_{A_{\mathrm{t}}}(a)$, the ν-parameter for the assumed distribution $f_{A_{\mathrm{a}}}(a)$ and the ν-parameter for the true distribution $f_{A_{\mathrm{t}}}(a)$, respectively. As argued in Section 3, $E\left[S|\mathbf{X}\right] = E\left[S|z(\mathbf{X})\right]$. Therefore, instead of computing Eq. (11) it is sufficient to compute the MSE by means of

$$\begin{aligned} &E_{z(\mathbf{x}),S}\left[|S - E_{\mathrm{a}}\left[S|z(\mathbf{X})\right]|^2\right] = \\ &\int\int g(z(\mathbf{X})) f_{z(\mathbf{X})}(z(\mathbf{x});\nu_{\mathrm{t}}) dz_{\Re}(\mathbf{X}) dz_{\Im}(\mathbf{X}). \end{aligned} \tag{12}$$

The function $g(z(\mathbf{X}))$ is given by

$$\begin{aligned} g(z(\mathbf{X})) &= |E_{\mathrm{a}}\left[S|z(\mathbf{X})\right]|^2 + E_{\mathrm{t}}\left[|S|^2|z(\mathbf{X})\right] \\ &- 2|E_{\mathrm{a}}\left[S|z(\mathbf{X})\right]||E_{\mathrm{t}}\left[S|z(\mathbf{X})\right]|. \end{aligned}$$

and the distribution $f_{z(\mathbf{X})}(z(\mathbf{x});\nu_{\mathrm{t}})$ in Eq. (12) is given by

$$\begin{aligned} &f_{z(\mathbf{X})}(z(\mathbf{x});\nu_{\mathrm{t}}) = \\ &\int_A \int_\Phi f_A(a;\nu_{\mathrm{t}}) f_\Phi(\phi) f_{z(\mathbf{X})|A,\Phi}(z(\mathbf{x})|a,\phi) d\phi da. \end{aligned}$$

Using the distributional assumptions in Section 2 and [14, 8.431.5,6.643.2,9.210.1 and 9.220.2] it then follows that

$$f_{z(\mathbf{X})}(z(\mathbf{x});\nu_{\mathrm{t}}) = \frac{e^{-\zeta_0}}{\pi\sigma_{V_0}^2}\left(\frac{\nu_{\mathrm{t}}}{\nu_{\mathrm{t}}+\xi_0}\right)^{\nu_{\mathrm{t}}} \mathcal{M}\left(\nu_{\mathrm{t}}, 1, \frac{\zeta_0\xi_0}{\nu_{\mathrm{t}}+\xi_0}\right).$$

with $\xi_0 = \sigma_S^2/\sigma_{V_0}^2$ and $\zeta_0 = |z(\mathbf{X})|^2/\sigma_{V_0}^2$ the *a priori* and *a posteriori* SNR with respect to $z(\mathbf{X})$, respectively, and $\sigma_{V_0}^2 = (\mathbf{d}^H\mathbf{\Sigma}^{-1}\mathbf{d})^{-1}$. Hence, for the special case $\nu_{\mathrm{t}} = 1$, $f_Z(z;\nu_{\mathrm{t}})$ is a complex Gaussian distribution with variance $\sigma_{V_0}^2 + \sigma_S^2$. Even though both the function $g(z(\mathbf{X}))$ and the distribution $f_{z(\mathbf{X})}(z(\mathbf{x});\nu_{\mathrm{t}})$ in Eq. (12) are known in closed form, the integrals in Eq. (12) cannot be solved analytically. Therefore, we compute the integral in Eq. (12) numerically based on the analytic expressions for $g(z(\mathbf{X}))$ and $f_{z(\mathbf{X})}(z(\mathbf{x});\nu_{\mathrm{t}})$. The resulting MSE is normalized by the MSE that would be obtained when using the true DFT-magnitude distribution, i.e.,

$$\mathrm{MSE}_{\mathrm{loss}} = 10\log_{10}\frac{E_{z(\mathbf{X}),S}\left[|S - E_{\mathrm{a}}\left[S|z(\mathbf{X})\right]|^2\right]}{E_{z(\mathbf{X}),S}\left[|S - E_{\mathrm{t}}\left[S|z(\mathbf{X})\right]|^2\right]}\ [\mathrm{dB}],$$

which expresses the loss in terms of MSE [dB] for making erroneous distributional assumptions.

In Fig. 1(a) and 1(b) we plot the $\mathrm{MSE}_{\mathrm{loss}}$ for the *a priori* SNR ξ_0 of 5 and 15 dB, respectively, as a function of ν_{t} for ν_{a}-values $\nu_{\mathrm{a}} \in \{0.1, 0.5, 1\}$. Obviously, we see that dependent on the true distribution $f_{A_{\mathrm{t}}}(a)$, specified by parameter ν_{t}, the $\mathrm{MSE}_{\mathrm{loss}}$ is minimized by the corresponding ν_{a}-value, i.e., when $\nu_{\mathrm{t}} = \nu_{\mathrm{a}}$. In practice, we know that speech DFT-coefficients can be best modelled with a ν-parameter in the range of 0.1 - 0.2, see e.g., [10][15]. From the MSE analysis in Fig. 1 we can make the following interesting observation. Imagine that the true distribution of speech DFT coefficients is super-Gaussian with $\nu_{\mathrm{t}} = 0.1$, but we assume that the speech DFT coefficients are Gaussian distributed, i.e., $\nu_{\mathrm{a}} = 1$, then a $\mathrm{MSE}_{\mathrm{loss}}$ of approximately 2.5 dB is obtained for the *a priori* SNRs of 5 and 15 dB. If we now consider the reverse situation, i.e., the true distribution of speech DFT coefficients is Gaussian with $\nu_{\mathrm{t}} = 1$ and the assumed distribution is super-Gaussian with $\nu_{\mathrm{a}} = 0.1$, then a $\mathrm{MSE}_{\mathrm{loss}}$ of approximately only 1.2 dB and 0.5 dB is obtained, for the *a priori* SNRs of 5

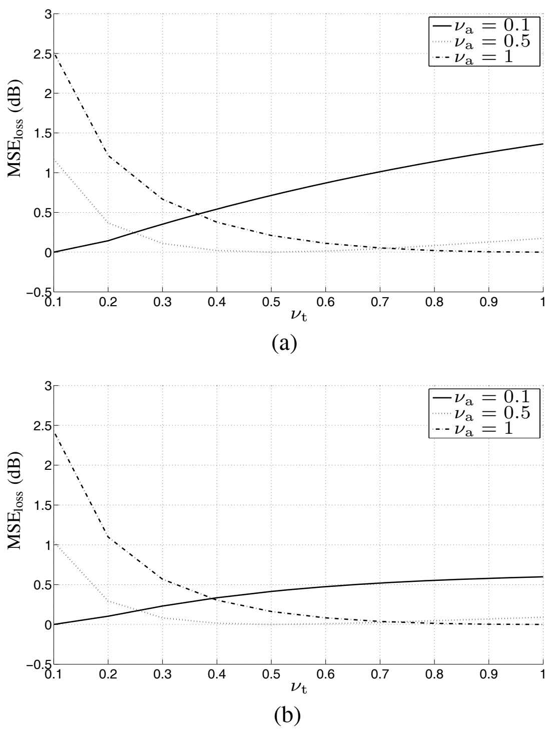

Figure 1: $\mathrm{MSE}_{\mathrm{loss}}$ as a function of ν_{t} for *a priori* SNR of (a) 5 dB (b) 15 dB.

and 15 dB, respectively. This suggests that if the true distribution of clean speech DFT coefficients were completely unknown, it is better in terms of MSE to use an estimator based on a super-Gaussian distribution with e.g., $\nu_{\mathrm{a}} = 0.1$, instead of assuming that the speech DFT coefficients are Gaussian distributed. Furthermore, this also suggests that the multi-channel MMSE estimator in Eq. (7)-(9) under the assumption that the speech DFT coefficients are super-Gaussian distributed is less sensitive to small deviations between the true super-Gaussian distribution (with a ν-parameter in the range of 0.1 - 0.2) and the assumed super-Gaussian distribution, than when the speech DFT coefficients would have been assumed Gaussian distributed, i.e., when using the MWF.

6. CONCLUSIONS

In this paper we discussed the robustness of multi-channel MMSE estimators under super-Gaussian speech priors with respect to deviations from the ideal steering vector. The multi-channel MMSE estimators can, under the given assumptions, always be decomposed into an MVDR beamformer and a generally non-linear single-channel post-processor applied to the MVDR-beamformer output. Given that there is a mismatch between the true and estimated steering vector, it can be concluded that the post-processor applied to the MVDR-beamformer output remains optimal. Further, we investigated the sensitivity of the multi-channel MMSE estimator when the assumed speech DFT distribution deviates from the true, but unknown, distribution. It can be concluded that the multi-channel MMSE estimator is less sensitive to small deviations between the true and assumed distribution when speech DFT coefficients are assumed to be super-Gaussian distributed instead of Gaussian distributed.

7. REFERENCES

[1] L. J. Griffiths, "A comparison of multidimensional Wiener and maximum-likelihood filters for antenna arrays," *Proceedings of the IEEE*, vol. 55, no. 11, pp. 2045–2047, Nov. 1967.

[2] 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. Aerospace and Electronic Systems*, vol. 8, no. 5, pp. 690–692, Sept. 1972.

[3] S. Doclo, "Multi-microphone noise reduction and dereverberation techniques for speech applications," Ph.D. dissertation, Katholieke Universiteit Leuven, 2003.

[4] J. C. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech and Language Processing*, vol. 14, no. 4, pp. 1218–1233, July 2006.

[5] S. Leukimmiatis, D. Dimitriadis, and P. Maragos, "An optimum microphone array post-filter for speech applications," in *Interspeech*, Sept. 2006, pp. 2142 – 2145.

[6] S. Gazor and W. Zhang, "Speech probability distribution," *IEEE Signal Processing Lett.*, vol. 10, no. 7, pp. 204–207, 2003.

[7] R. Martin, "Speech enhancement based on minimum mean-square error estimation and supergaussian priors," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 5, pp. 845–856, Sept. 2005.

[8] T. Lotter and P. Vary, "Speech enhancement by MAP spectral amplitude estimation using a super-Gaussian speech model," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 7, pp. 1110–1126, 2005.

[9] R. C. Hendriks, R. Heusdens, U. Kjems, and J. Jensen, "On optimal multi-channel mean-squared error estimators for speech enhancement," *IEEE Signal Processing Letters*, 2009.

[10] 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 and Language Processing*, vol. 15, no. 6, pp. 1741 – 1752, August 2007.

[11] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. New York: Wiley, 2001.

[12] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Philadelphia: SIAM, 2001.

[13] H. L. van Trees, *Detection, Estimation and Modulation Theory*. John Wiley and Sons, 1968, vol. 1.

[14] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, 6th ed. New York: Academic, 2000.

[15] 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.