MULTI-CHANNEL PSD ESTIMATORS FOR SPEECH DEREVERBERATION – A THEORETICAL AND EXPERIMENTAL COMPARISON

Adam Kuklasiński[*,†], *Simon Doclo*[‡], *Timo Gerkmann*[‡], *Søren Holdt Jensen*[†], *Jesper Jensen*[*,†]

[*]Oticon A/S, 2765 Smørum, Denmark
[†]Aalborg University, Department of Electronic Systems, 9220 Aalborg, Denmark
[‡]University of Oldenburg, Department of Medical Physics and Acoustics, and Cluster of Excellence Hearing4all, Oldenburg, Germany

ABSTRACT

In this paper we perform an extensive theoretical and experimental comparison of two recently proposed multi-channel speech dereverberation algorithms. Both of them are based on the multi-channel Wiener filter but they use different estimators of the speech and reverberation power spectral densities (PSDs). We first derive closed-form expressions for the mean square error (MSE) of both PSD estimators and then show that one estimator – previously used for speech dereverberation by the authors – always yields a better MSE. Only in the case of a two microphone array or for special spatial distributions of the interference both estimators yield the same MSE. The theoretically derived MSE values are in good agreement with numerical simulation results and with instrumental speech quality measures in a realistic speech dereverberation task for binaural hearing aids.

Index Terms— PSD estimation, maximum likelihood estimation, multi-channel Wiener filter, speech dereverberation, isotropic.

1. INTRODUCTION

Background noise and reverberation may have a detrimental effect on speech quality and intelligibility [1]. Consequently, speech denoising and dereverberation algorithms are of interest in many applications, e.g. hearing aids, mobile phones, etc. Many of these devices contain multiple microphones, which enables the use of spatial filtering algorithms such as the Multi-channel Wiener Filter (MWF) [2,3]. Under a set of commonly made assumptions the MWF is an optimal estimator of the speech signal in the Minimum Mean Square Error (MMSE) sense [2]. However, in order to obtain its theoretical performance the MWF requires knowledge of the (cross-) Power Spectral Density (PSD) matrices of the target (speech) and interference (noise, reverberation) signal components. These are usually unknown and have to be estimated from the noisy and reverberant microphone signals. In practice, the performance of the resulting MWF depends on the accuracy of the used PSD estimation scheme.

In this paper we compare two multi-channel speech dereverberation algorithms recently proposed in [4] and [5]. Both algorithms are based on the MWF and use the assumption that the reverberant sound field is cylindrically isotropic. The PSD estimators used in [4] and [5] are both derived using the Maximum (ML) methodology but use different statistical assumptions, and therefore yield different formulas and results.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement Nº ITN-GA-2012-316969. More information about the project can be found on the website: /www.dreams-itn.eu/.

In order to perform a theoretical comparison of the two PSD estimation schemes we first derive analytical expressions for their Mean Square Error (MSE). This allows us to show that the PSD estimators used in [4] achieve lower or equal MSE compared to the PSD estimators in [5]. We also derive the conditions under which the two PSD estimation schemes yield the same MSE. We verify these theoretical results using numerical simulations.

Finally, we evaluate the speech dereverberation performance of the MWFs from [4] and [5] in a simulation of binaural hearing aids in realistic reverberant conditions. The results of the experiment show that the algorithm from [4] outperforms [5] in terms of objective performance measures such as Frequency-Weighted Segmental SNR (FWSegSNR) [6] and Perceptual Evaluation of Speech Quality (PESQ) [7].

2. SIGNAL MODEL AND ASSUMPTIONS

The signal model and assumptions that are used in the speech dereverberation algorithms proposed in [4] and [5] share many characteristics. Both algorithms operate on Short Time Fourier Transform (STFT) coefficients $y_m(k,n)$ which are computed from the time domain signals $y_m(t)$ of M microphones:

$$y_m(k,n) = \sum_{t=0}^{T-1} y_m(t+nD)w(t)e^{-2\pi i k\frac{t}{T}}, \quad m = 1,\dots,M,$$

where n is the time frame index and k is the frequency bin index. The STFT order is denoted by T, the filterbank decimation factor by D, and $w(t)$ is the analysis window function. The algorithms from [4] and [5] process the individual frequency bins independently of each other. This enables us to omit the index k without loss of generality. For notational convenience the STFT coefficients corresponding to all microphones are stacked in a vector as: $\mathbf{y}(n) = [\, y_1(n) \dots\ y_M(n)\,]^T$.

The algorithms from [4] and [5] employ an additive model of the reverberant speech signal:

$$\mathbf{y}(n) = \mathbf{s}(n) + \mathbf{v}(n) \overbrace{+\mathbf{x}(n)}^{\text{only in [5]}}, \tag{1}$$

where $\mathbf{s}(n)$ denotes the direct-path speech component and $\mathbf{v}(n)$ denotes the reverberation component of the microphone signal. The algorithm from [5] allows for an additional noise term $\mathbf{x}(n)$, whose cross-PSD matrix must be known. In this study, for mathematical convenience, we assume that this additional noise component is equal to zero. This corresponds to an assumption that $\mathbf{x}(n)$ is negligible compared to the reverberation component, which may be valid

in some situations. It is assumed that the vectors $\mathbf{s}(n)$ and $\mathbf{v}(n)$ are statistically independent across time frames and frequency bins.

The algorithms from [4] and [5] aim to estimate the direct-path speech signal component $s(n)$ at a certain reference position, e.g. one of the microphones. Because the speech is assumed to be generated by a point source, the vector $\mathbf{s}(n)$ may be written as the product of $s(n)$ and a vector of Relative Transfer Functions (RTFs) $\mathbf{d}$ [8]:

$$\mathbf{y}(n) = s(n)\mathbf{d} + \mathbf{v}(n).$$

The elements of $\mathbf{d}$ correspond to the transfer functions of the direct-path speech from the chosen reference position to all microphones. In [4] and [5] the RTF vector $\mathbf{d}$ is assumed to be known.

In both algorithms the focus is on reducing the late part of the reverberation, which is assumed to be uncorrelated with the direct-path speech. Hence, the cross-PSD matrix of $\mathbf{y}(n)$ can be modeled as the sum of the speech and the reverberation cross-PSD matrices:

$$\mathbf{\Phi}_{\mathbf{y}}(n) = E\big[\mathbf{y}(n)\mathbf{y}^H(n)\big] = \mathbf{\Phi}_{\mathbf{s}}(n) + \mathbf{\Phi}_{\mathbf{v}}(n).$$

where $E[\,]$ denotes the expectation operator. Because of the assumption that the speech is generated by a point source, $\mathbf{\Phi}_{\mathbf{s}}(n)$ is modeled as a rank-one matrix and can be written in terms of the scalar PSD $\phi_s(n)$ of the direct-path speech at the reference position and the RTF vector $\mathbf{d}$: $\phi_s(n)\mathbf{d}\mathbf{d}^H$. Similarly, matrix $\mathbf{\Phi}_{\mathbf{v}}(n)$ may be written as a product of the scalar PSD $\phi_v(n)$ of the reverberation at the reference position, and the cross-PSD matrix $\mathbf{\Gamma}_{\mathbf{v}}$ of the reverberation normalized by $\phi_v(n)$:

$$\mathbf{\Phi}_{\mathbf{y}}(n) = \phi_s(n)\mathbf{d}\mathbf{d}^H + \phi_v(n)\mathbf{\Gamma}_{\mathbf{v}}, \tag{2}$$

Due to the assumption of cylindrical isotropy of the reverberant sound field made in both [4] and [5], the matrix $\mathbf{\Gamma}_{\mathbf{v}}$ is assumed to be constant, full-rank, and known. For free-field microphone arrays, $\mathbf{\Gamma}_{\mathbf{v}}$ can even be calculated analytically using information on microphone array geometry (as in [5]). Alternatively, e.g. for hearing aid applications, $\mathbf{\Gamma}_{\mathbf{v}}$ may be estimated from measurements using the actual microphone array in a (possibly simulated) isotropic sound field (as in [4]). While the vector $\mathbf{d}$ and the matrix $\mathbf{\Gamma}_{\mathbf{v}}$ are assumed to be known and constant, the PSDs $\phi_s(n)$ and $\phi_v(n)$ are unknown and time-varying because of the non-stationarity of $\mathbf{s}(n)$ and $\mathbf{v}(n)$.

3. MULTI-CHANNEL WIENER FILTER

The algorithms from [4] and [5] are both based on the Multi-channel Wiener Filter (MWF) [2, 3]. It is well-known that the MWF is an MMSE-optimal estimator of the target speech $s(n)$ if the input signal components $\mathbf{s}(n)$ and $\mathbf{v}(n)$ are normally distributed, or alternatively, if the search is limited to linear estimators. Because of the rank-one assumption on $\mathbf{\Phi}_{\mathbf{s}}(n)$, the MWF may be factorized into an MVDR beamformer $\mathbf{w}_{\mathrm{mvdr}}$ and a single-channel Wiener filter $g_{\mathrm{sc}}(n)$ [2]:

$$\begin{aligned} \hat{s}(n) &= \mathbf{w}_{\mathrm{mwf}}^H(n)\mathbf{y}(n), \\ \mathbf{w}_{\mathrm{mwf}}(n) &= \underbrace{\left[\frac{\phi_{s_o}(n)}{\phi_{s_o}(n)+\phi_{v_o}(n)}\right]}_{g_{\mathrm{sc}}(n)} \underbrace{\frac{\mathbf{\Gamma}_{\mathbf{v}}^{-1}\mathbf{d}}{\mathbf{d}^H\mathbf{\Gamma}_{\mathbf{v}}^{-1}\mathbf{d}}}_{\mathbf{w}_{\mathrm{mvdr}}}, \end{aligned} \tag{3}$$

where $\phi_{s_o}(n)$ and $\phi_{v_o}(n)$ are the PSDs of the direct-path speech and reverberation at the output of the MVDR beamformer, i.e.: $\phi_{s_o}(n) = \phi_s(n)$, and $\phi_{v_o}(n) = \mathbf{w}_{\mathrm{mvdr}}^H\phi_v(n)\mathbf{\Gamma}_{\mathbf{v}}\mathbf{w}_{\mathrm{mvdr}}$. For the signal model described in Sec. 2 the vector $\mathbf{w}_{\mathrm{mvdr}}$ is constant and is readily calculated from $\mathbf{d}$ and $\mathbf{\Gamma}_{\mathbf{v}}$.

4. POWER SPECTRAL DENSITY ESTIMATION

The main difference between the algorithms from [4] and [5] is the method used to estimate the unknown PSDs of the direct-path speech $\phi_s(n)$ and of the reverberation $\phi_v(n)$. In this section, we briefly review these two PSD estimation schemes.

4.1. Algorithm [4] by Kuklasiński et al.

The PSD estimators used in [4] are based on the assumption that the STFT coefficients of the signal components are circularly-symmetric complex Gaussian distributed, i.e.:

$$\mathbf{s}(n) \sim \mathcal{CN}\big(\mathbf{0}, \mathbf{\Phi}_{\mathbf{s}}(n)\big), \qquad \mathbf{v}(n) \sim \mathcal{CN}\big(\mathbf{0}, \mathbf{\Phi}_{\mathbf{v}}(n)\big).$$

The above distributions can be used to construct a likelihood function, compute its partial derivatives, and ultimately, derive a pair of joint Maximum Likelihood Estimators (MLEs) of $\phi_s(n)$ and $\phi_v(n)$. Several formulations of these estimators are available in the literature [9, 10], but in [4] the one from [9] has been used:

$$\hat{\phi}_{v,[4]}(n) = \frac{1}{M-1}\,\mathrm{tr}\Big[\big(\mathbf{I} - \mathbf{d}\mathbf{w}_{\mathrm{mvdr}}^H\big)\hat{\mathbf{\Phi}}_{\mathbf{y}}(n)\,\mathbf{\Gamma}_{\mathbf{v}}^{-1}\Big], \tag{4a}$$

$$\hat{\phi}_{s,[4]}(n) = \mathbf{w}_{\mathrm{mvdr}}^H\Big[\hat{\mathbf{\Phi}}_{\mathbf{y}}(n) - \hat{\phi}_{v,[4]}(n)\,\mathbf{\Gamma}_{\mathbf{v}}\Big]\mathbf{w}_{\mathrm{mvdr}}, \tag{4b}$$

where $\mathrm{tr}[\,]$ denotes the matrix trace, $\hat{\mathbf{\Phi}}_{\mathbf{y}}(n)$ denotes the estimate of the cross-PSD matrix of the input signal:

$$\hat{\mathbf{\Phi}}_{\mathbf{y}}(n) = \frac{1}{L}\sum_{l=0}^{L-1}\mathbf{y}(n-l)\mathbf{y}^H(n-l), \tag{5}$$

and where the PSDs $\phi_s(n)$ and $\phi_v(n)$ are assumed to be constant across the L averaged STFT frames.

4.2. Algorithm [5] by Braun and Habets

Similarly to [4], in [5] the reverberation PSD estimator is derived using the ML methodology. However, the likelihood function used in the derivation is based on a different statistical assumption than in [4], resulting in a different estimator.

Specifically, the reverberation PSD estimator used in [5] is derived by first defining a blocking matrix $\mathbf{B} \in \mathbb{C}_{M\times(M-1)}$ which represents a set of $M-1$ target-canceling beamformers. In [5] it is computed according to the method used in [10]:

$$\big[\mathbf{B}\ \mathbf{b}\big] = \mathbf{A}, \quad \mathbf{A} = \mathbf{I} - \mathbf{d}(\mathbf{d}^H\mathbf{d})^{-1}\mathbf{d}^H.$$

Next, an error matrix $\mathbf{\Phi}_{\mathrm{err}}(n)$ is defined as:

$$\mathbf{\Phi}_{\mathrm{err}}(n) = \hat{\tilde{\mathbf{\Phi}}}_{\mathbf{y}}(n) - \phi_v(n)\tilde{\mathbf{\Gamma}}_{\mathbf{v}}, \tag{6}$$

with $\tilde{\mathbf{\Gamma}}_{\mathbf{v}} = \mathbf{B}^H\mathbf{\Gamma}_{\mathbf{v}}\mathbf{B}$ and

$$\hat{\tilde{\mathbf{\Phi}}}_{\mathbf{y}}(n) = \mathbf{B}^H\hat{\mathbf{\Phi}}_{\mathbf{y}}(n)\mathbf{B}. \tag{7}$$

The matrix $\hat{\tilde{\mathbf{\Phi}}}_{\mathbf{y}}(n)$ is the estimate of the cross-PSD matrix of the blocked input signal $\tilde{\mathbf{y}}(n) = \mathbf{B}^H\mathbf{y}(n)$. Because $\mathbf{B}^H\mathbf{s}(n) = \mathbf{0}$, $\hat{\tilde{\mathbf{\Phi}}}_{\mathbf{y}}(n)$ is equivalently the estimate of the cross-PSD matrix of the blocked reverberation signal component $\mathbf{B}^H\mathbf{v}(n)$ (cf. (1)). Hence, the matrix $\mathbf{\Phi}_{\mathrm{err}}(n)$ in (6) can be interpreted as the error between the blocked reverberation cross-PSD matrix $\phi_v(n)\tilde{\mathbf{\Gamma}}_{\mathbf{v}}$ (cf. (2)) and its estimate $\hat{\tilde{\mathbf{\Phi}}}_{\mathbf{y}}(n)$. In [5] the elements of $\mathbf{\Phi}_{\mathrm{err}}(n)$ are modeled as independent circularly-symmetric complex Gaussian random variables

of equal variance. This assumption is used to construct a likelihood function from which an MLE of $\phi_v(n)$ is calculated as [5]:

$$\hat{\phi}_{v,[5]}(n) = \mathrm{tr}\Big[\tilde{\mathbf{\Gamma}}_{\mathbf{v}}\hat{\tilde{\mathbf{\Phi}}}_{\mathbf{y}}(n)\Big]\mathrm{tr}\Big[\tilde{\mathbf{\Gamma}}_{\mathbf{v}}^2\Big]^{-1}. \tag{8a}$$

The corresponding estimator of $\phi_s(n)$ is derived without the use of the ML methodology, but coincidentally has the same form as the MLE used in [4] (4b):

$$\hat{\phi}_{s,[5]}(n) = \mathbf{w}_{\mathrm{mvdr}}^H\Big[\hat{\mathbf{\Phi}}_{\mathbf{y}}(n) - \hat{\phi}_{v,[5]}(n)\,\mathbf{\Gamma}_{\mathbf{v}}\Big]\mathbf{w}_{\mathrm{mvdr}}. \tag{8b}$$

5. ANALYTICAL EVALUATION

In this section we analytically derive the MSE of the reverberation PSD estimator from [5] and compare it to the MSE of the reverberation PSD estimator from [4]. Differences between the direct-path speech PSD estimators from [4] and [5] are exclusively due to the different reverberation PSD estimators used in (4b) and (8b). Therefore, relations between the MSEs of the direct-path speech PSD estimators are completely determined by and are analogous to the relations between the MSEs of the reverberation PSD estimators.

We start by noting that the PSD estimators from [4] and [5] are unbiased (without proof):

$$E\big[\hat{\phi}_{p,r}(n)\big] = \phi_p(n), \quad p \in \{s, v\}, \quad r \in \{[4], [5]\}. \tag{9}$$

Hence, the MSEs of these estimators are identical to their variances.

The variance of the direct-path speech PSD estimator from [4] can be shown to be equal to the corresponding asymptotic Cramér-Rao Lower Bound (CRLB) which is equal to [11]:

$$\begin{aligned}\mathrm{var}\big(\hat{\phi}_{s,[4]}(n)\big) &= \mathrm{CRLB}\big(\hat{\phi}_s(n)\big)\\ &= \phi_s^2(n)\frac{1}{L}\left[\left(\frac{1+\xi(n)}{\xi(n)}\right)^2 + \frac{1}{M-1}\frac{1}{\xi^2(n)}\right],\end{aligned} \tag{10}$$

where $\xi(n) = \phi_{s_o}(n)/\phi_{v_o}(n)$ is the SNR at the output of the MVDR beamformer. From [11] it also follows that the variance of the reverberation PSD estimator from [4] is equal to the respective CRLB, and that this CRLB is equal to:

$$\mathrm{var}\big(\hat{\phi}_{v,[4]}(n)\big) = \mathrm{CRLB}\big(\hat{\phi}_v(n)\big) = \phi_v^2(n)\frac{1}{L}\frac{1}{M-1}. \tag{11}$$

We derive the variance of the reverberation PSD estimator from [5] by using (8a), (7) and (5) and moving the deterministic factors outside the variance operator:

$$\mathrm{var}\big(\hat{\phi}_{v,[5]}(n)\big) = \mathrm{tr}\Big[\tilde{\mathbf{\Gamma}}_{\mathbf{v}}^2\Big]^{-2}\frac{1}{L}\,\mathrm{var}\Big(\mathrm{tr}\Big[\tilde{\mathbf{y}}^H(n)\tilde{\mathbf{\Gamma}}_{\mathbf{v}}\tilde{\mathbf{y}}(n)\Big]\Big),$$

The trace operator inside the variance operator may now be omitted because its argument has been reduced to a quadratic form (a scalar). The variance of such quadratic forms in circularly-symmetric complex Gaussian random vectors is given by [12, p. 513, eq. (15.30)]:

$$\mathrm{var}\big(\mathbf{a}^H\mathbf{Z}\mathbf{a}\big) = \mathrm{tr}\big(\mathbf{\Phi}_{\mathbf{a}}\mathbf{Z}\mathbf{\Phi}_{\mathbf{a}}\mathbf{Z}\big), \text{ where } \mathbf{a} \sim \mathcal{CN}\big(\mathbf{0}, \mathbf{\Phi}_{\mathbf{a}}\big). \tag{12}$$

Using (12) and the fact that $\tilde{\mathbf{y}}(n) \sim \mathcal{CN}\big(\mathbf{0}, \phi_v(n)\tilde{\mathbf{\Gamma}}_{\mathbf{v}}\big)$, we obtain:

$$\mathrm{var}\big(\hat{\phi}_{v,[5]}(n)\big) = \phi_v^2(n)\frac{1}{L}\,\mathrm{tr}\Big[\tilde{\mathbf{\Gamma}}_{\mathbf{v}}^4\Big]\,\mathrm{tr}\Big[\tilde{\mathbf{\Gamma}}_{\mathbf{v}}^2\Big]^{-2}. \tag{13}$$

Before comparing (11) and (13), we transform (13) into a more convenient form. Let $\tilde{\mathbf{\Gamma}}_{\mathbf{v}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ denote the eigenvalue decomposition of the positive-definite Hermitian matrix $\tilde{\mathbf{\Gamma}}_{\mathbf{v}}$, where $\mathbf{\Lambda}$ is a diagonal matrix containing the $M-1$ positive eigenvalues $\lambda_1, \ldots, \lambda_{M-1}$ of $\tilde{\mathbf{\Gamma}}_{\mathbf{v}}$. Using the facts that $\mathrm{tr}(\tilde{\mathbf{\Gamma}}_{\mathbf{v}}) = \sum_{m=1}^{M-1}\lambda_m$, $\tilde{\mathbf{\Gamma}}_{\mathbf{v}}^p = \mathbf{V}\mathbf{\Lambda}^p\mathbf{V}^H$, and defining $\gamma_m = \lambda_m^2$, (13) may be written as:

$$\mathrm{var}\big(\hat{\phi}_{v,[5]}(n)\big) = \phi_v^2(n)\frac{1}{L}\,\frac{\sum_{m=1}^{M-1}\gamma_m^2}{\big(\sum_{m=1}^{M-1}\gamma_m\big)^2}. \tag{14}$$

If we denote the average of the squared eigenvalues γ_m by $\bar{\gamma}$, and the sample variance of these squared eigenvalues around $\bar{\gamma}$ by $\tilde{\gamma}^2$,

$$\bar{\gamma} = \frac{1}{M-1}\sum_{m=1}^{M-1}\gamma_m, \quad \tilde{\gamma}^2 = \left(\frac{1}{M-1}\sum_{m=1}^{M-1}\gamma_m^2\right) - \bar{\gamma}^2,$$

then we can rewrite (13) as:

$$\mathrm{var}\big(\hat{\phi}_{v,[5]}(n)\big) = \phi_v^2(n)\frac{1}{L}\frac{1}{M-1}\left(1 + \frac{\tilde{\gamma}^2}{\bar{\gamma}^2}\right). \tag{15}$$

Comparing (15) and (11) we can now deduce, that the MSE of $\hat{\phi}_{v,[5]}(n)$ can be either greater or equal to the MSE of $\hat{\phi}_{v,[4]}(n)$ (and the CRLB), but can never be lower. The MSEs of these two estimators are equal only when the eigenvalues of $\tilde{\mathbf{\Gamma}}_{\mathbf{v}}$ are all equal (i.e. when $\tilde{\gamma}^2 = 0$). Since $\tilde{\mathbf{\Gamma}}_{\mathbf{v}}$ is Hermitian, it follows that for this special case to occur, $\tilde{\mathbf{\Gamma}}_{\mathbf{v}}$ must be a scaled identity matrix [13]. In all other cases, the reverberation PSD estimator from [4] outperforms the one from [5]. An important observation is that for $M = 2$ the matrix $\tilde{\mathbf{\Gamma}}_{\mathbf{v}}$ reduces to a scalar, such that $\tilde{\gamma}^2$ is always equal to zero. It follows that for $M = 2$ the reverberation PSD estimators from [4] and [5] achieve the same MSE under all possible conditions.

We can also compute the upper bound of the variance of the reverberation PSD estimator from [5]. The ratio $\tilde{\gamma}^2/\bar{\gamma}^2$ in (15) is maximal when all but one eigenvalue tend to zero (all energy is concentrated in a single eigenvalue). This may occur when the interference is dominated by one directional component. For such interferences the variance (and MSE) of $\hat{\phi}_{v,[5]}(n)$ equals:

$$\max_{\tilde{\mathbf{\Gamma}}_{\mathbf{v}}}\mathrm{var}\big(\hat{\phi}_{v,[5]}(n)\big) = \phi_v^2(n)\,\frac{1}{L} \tag{16}$$

i.e. is $M-1$ times larger than that of $\hat{\phi}_{v,[4]}(n)$.

6. EXPERIMENTAL EVALUATION

We first confirm our theoretical results using a series of numerical simulations (Sec. 6.1). Additionally, we evaluate the MWF algorithms from [4] and [5] in a speech dereverberation experiment (Sec. 6.2). In both experiments the microphone array is composed of a pair of Oticon Epoq behind-the-ear hearing aids [14], each containing two microphones (i.e. $M = 4$). We have measured the RTF vectors $\mathbf{d}$ and the matrices $\mathbf{\Gamma}_{\mathbf{v}}$ in an anechoic chamber with the hearing-aids placed on a Head And Torso acoustic Simulator (HATS). The reference position for calculating $\mathbf{d}$ and $\mathbf{\Gamma}_{\mathbf{v}}$ was chosen as one of the microphones ($m = 1$), such that the corresponding elements of $\mathbf{d}$ and $\mathbf{\Gamma}_{\mathbf{v}}$ were equal to one. We used the RTF vector measured for the source position directly in front of the HATS.

6.1. Experiment 1: MSE of PSD estimation

In order to verify the theoretical results of Sec. 5, we have conducted a number of iterations of a numerical simulation. In each iteration a test signal $\mathbf{y}(n)$ was generated using $N = 25000$ pseudo-random STFT vectors drawn from a circularly-symmetric multivariate complex Gaussian distribution. The covariance matrix of this distribution

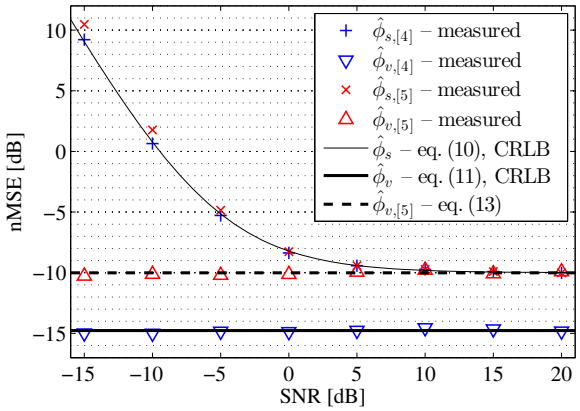

Fig. 1. Normalized MSE of the reverberation and direct-path speech PSD estimators from [4] and [5], as a function of the input SNR, measured numerically and compared to the theoretical values. ($M = 4$, $f = 1\,\text{kHz}$, $L = 10$)

was modeled according to (2), using the measured RTF vector $\mathbf{d}$ and matrix $\mathbf{\Gamma_v}$ for the STFT frequency bin corresponding to 1 kHz. In each iteration $\phi_s(n)$ and $\phi_v(n)$ were set to correspond to different input SNRs between -15 and 20 dB at the reference microphone.

Next, the PSD estimators from [4] and [5] were used to estimate $\phi_s(n)$ and $\phi_v(n)$ of the test signals. The averaging length in (5) was set to $L = 10$ frames. Because the true values of the PSDs were known, it was possible to compute the MSE achieved by each of the estimators under each of the simulated SNRs. To facilitate the comparison of the obtained results, we normalized the measured MSEs by the square of the parameter of interest:

$$\text{nMSE}(\hat{\phi}_{p,r}) = \frac{\text{MSE}(\hat{\phi}_{p,r})}{\phi_p^2} = \frac{1}{N-L+1}\sum_{n=L}^{N}\frac{(\hat{\phi}_{p,r}(n)-\phi_p)^2}{\phi_p^2},$$

with p and r defined as in (9).

The results of this experiment are presented in Fig. 1. For comparison, the analytically derived nMSEs formulated in (10), (11), and (13) are also included in the plot. The results of the numerical simulation closely agree with the theoretical formulas. The MSE achieved by the direct-path speech PSD estimator from [5] is close to, but greater than the MSE achieved by the estimator from [4]. It can also be observed that in the particular example of the simulated binaural hearing aid configuration of the microphone array, the advantage of using (4a) over (8a) for estimating the reverberation PSD is approximately 5 dB MSE for all input SNRs. Moreover, the nMSE achieved by the reverberation PSD estimator from [5] is close to the upper bound derived in (16), which for $L = 10$ equals -10 dB nMSE. This indicates, that the reverberation PSD estimator from [5] is not optimally suited for the simulated acoustic scenario.

6.2. Experiment 2: speech dereverberation performance

In order to evaluate the influence of the different PSD estimators on the MWF performance, we conducted a second simulation experiment analogous to the one presented in [4]. In this experiment the test signals were synthesized by convolving TIMIT speech sentences [15] with six different multi-channel impulse responses. Five of them were measured in real rooms using a similar microphone array as for measuring $\mathbf{d}$ and $\mathbf{\Gamma_v}$. The sixth multi-channel impulse response (denoted "Isotropic") was synthesized to simulate an ideal

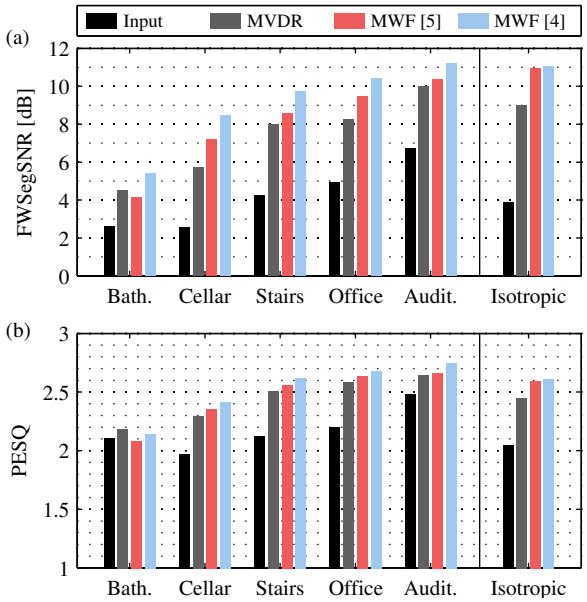

Fig. 2. (a) FWSegSNR and (b) PESQ scores of the algorithms from [4] and [5] (denoted "MWF"). The scores computed from the unprocessed signal $y_1(n)$ ("Input"), and the output of the MVDR beamformer $\mathbf{w}_{\text{mvdr}}^H\mathbf{y}(n)$ ("MVDR") are also included.

cylindrically isotropic reverberant sound field. In the present study, the same room impulse responses and the same values of the non-critical simulation parameters have been used as in [4], where their detailed description may be found.

The algorithms from [4] and [5] were used to dereverberate the test signals and their performance was evaluated using the Frequency-Weighted Segmental SNR (FWSegSNR) [6] and Perceptual Evaluation of Speech Quality (PESQ) [7] objective measures. The results of this evaluation are presented in Fig. 2. In can be observed, that the lower MSE of the PSD estimators used in the algorithm from [4] results in a better speech dereverberation performance as measured using FWSegSNR and PESQ. Although the difference is small in the "Isotropic" condition, the advantage of using [4] over [5] increases in all realistic reverberation conditions simulated in this experiment. This suggests, that the speech dereverberation algorithm proposed in [4] may be more robust to deviations from the assumed cylindrical isotropy of the reverberation, which necessarily occur in real rooms.

7. CONCLUSION

In this paper we have compared two similar speech dereverberation algorithms proposed in [4] and [5]. Theoretical analysis of the direct-path speech and reverberation PSD estimators used in both algorithms revealed that for microphone numbers greater than two, the estimators used in [4] perform better than the ones used in [5] in almost all conditions. These theoretical results were confirmed in a numerical simulation.

The speech dereverberation performance of the algorithms from [4] and [5] in a four microphone binaural hearing aid configuration was measured in realistic reverberation conditions. It is found that the dereverberation algorithm from [4] outperforms [5] in terms of the FWSegSNR and PESQ objective performance measures.

8. REFERENCES

[1] A. K. Nabelek and J. M. Pickett, "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," *J. Speech, Lang. Hearing Res.*, vol. 17, no. 4, pp. 724–739, 1974.

[2] S. Doclo et al., "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. J. Ray Liu, Eds., pp. 269–302. Wiley, 2008.

[3] S. Doclo et al., "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Commun.*, vol. 49, no. 7-8, pp. 636–656, 2007.

[4] A. Kuklasiński et al., "Maximum likelihood based multichannel isotropic reverberation reduction for hearing aids," in *Proc. 22nd Eur. Signal Process. Conf. (EUSIPCO)*, Lisbon, Portugal, 2014, pp. 61–65.

[5] S. Braun and E. A. P. Habets, "Dereverberation in noisy environments using reference signals and a maximum likelihood estimator," in *Proc. 21st Eur. Signal Process. Conf. (EUSIPCO)*, Marrakech, Morocco, 2013, pp. 1–5.

[6] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, pp. 229–238, Jan. 2008.

[7] "Perceptual evaluation of speech quality: an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *ITU-T Rec. P. 862*, 2001.

[8] S. Gannot et al., "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, pp. 1614–1626, Aug. 2001.

[9] H. Ye and R. D. DeGroat, "Maximum likelihood DOA estimation and asymptotic Cramér-Rao bounds for additive unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, pp. 938–949, Apr. 1995.

[10] U. Kjems and J. Jensen, "Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Bucharest, Romania, 2012, pp. 295–299.

[11] J. Jensen and M. S. Pedersen, "Analysis of beamformer-directed single-channel noise reduction system for hearing aid applications," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Brisbane, Australia, 2015.

[12] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*, Prentice Hall Signal Processing Series. Prentice-Hall PTR, 1993.

[13] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[14] Oticon A/S, "Oticon Epoq brochure," `/www.oticon.com/support/hearing-aids/downloads/legacy-products/~asset/cache.ashx?id=2727&type=14`, [Online; accessed 19-Jan-2015].

[15] J. S. Garofolo et al., *DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM*, NIST, 1993.