# LATE REVERBERANT POWER SPECTRAL DENSITY ESTIMATION BASED ON AN EIGENVALUE DECOMPOSITION

*Ina Kodrasi, Simon Doclo*

University of Oldenburg, Department of Medical Physics and Acoustics, and Cluster of Excellence Hearing4All, Oldenburg, Germany
{ina.kodrasi,simon.doclo}@uni-oldenburg.de

## ABSTRACT

Multi-channel methods for estimating the late reverberant power spectral density (PSD) rely on an estimate of the direction of arrival (DOA) of the speech source or of the relative early transfer functions (RETFs) of the target signal from a reference microphone to all microphones. The DOA and the RETFs may be difficult to estimate accurately, particularly in highly reverberant and noisy scenarios. In this paper we propose a novel multi-channel method to estimate the late reverberant PSD which does not require estimates of the DOA or RETFs. The late reverberation is modeled as an isotropic sound field and the late reverberant PSD is estimated based on the eigenvalues of the prewhitened received signal PSD matrix. Experimental results demonstrate the advantages of using the proposed estimator in a multi-channel Wiener filter for speech dereverberation, outperforming a recently proposed maximum likelihood estimator both when the DOA is perfectly estimated as well as in the presence of DOA estimation errors.

***Index Terms***— speech dereverberation, MWF, late reverberant PSD, EVD, prewhitening

## 1. INTRODUCTION

In many speech communication applications the received microphone signals are corrupted by reverberation, typically leading to decreased speech quality and intelligibility [1–3] and performance deterioration in speech recognition systems [4, 5]. Since late reverberation is the major cause of speech quality and intelligibility degradation, effective enhancement techniques that reduce the late reverberation are required. In the last decades many single- and multi-channel dereverberation techniques have been proposed [6], with multi-channel techniques being generally preferred since they exploit both the spectro-temporal and the spatial characteristics of the received microphone signals. Many such techniques require an estimate of the late reverberant power spectral density (PSD), e.g., [7–9]. The late reverberant PSD can be estimated using single- or multi-channel estimators, with multi-channel estimators shown to yield a higher PSD estimation accuracy [10].

Dual-channel coherence-based estimators are proposed in e.g., [11, 12], which however exploit only two microphones. A multi-channel maximum likelihood (ML) estimator is proposed in [7], where the late reverberation is modeled as an isotropic sound field and the late reverberant PSD is estimated from a set of reference signals at the output of a blocking matrix. In [8] the use of a blocking matrix is circumvented and an ML estimate of the late reverberant PSD is derived from the received microphone signals. In [13], it is theoretically and experimentally validated that the ML estimator proposed in [8] yields a higher PSD estimation accuracy than the ML estimator proposed in [7]. While a noise-free scenario is assumed in [8], late reverberant PSD estimators for noisy scenarios are proposed in [7, 14–17]. All multi-channel late reverberant PSD estimators in [7, 8, 14–17] require an estimate of the direction of arrival (DOA) of the speech source or of the relative early transfer functions (RETFs) of the target signal from a reference microphone to all microphones. The DOA and the RETFs may be difficult to estimate accurately, particularly in highly reverberant and noisy scenarios. As is experimentally validated in [12, 18], DOA estimation errors degrade the PSD estimation accuracy, yielding as a result a degradation in the dereverberation performance of the used speech enhancement system.

In this paper a novel multi-channel late reverberant PSD estimator is proposed, which does not require knowledge of the DOA or RETFs. The late reverberation is modeled as an isotropic sound field and the late reverberant PSD is estimated based on the eigenvalues of the prewhitened received signal PSD matrix. Experimental results demonstrate the advantages of using the proposed estimator in a multi-channel Wiener filter (MWF) for speech dereverberation, outperforming the ML estimator in [8] both when the DOA is perfectly estimated as well as in the presence of DOA estimation errors.

## 2. CONFIGURATION AND NOTATION

Consider a reverberant acoustic system with a single speech source and $M \geq 2$ microphones, as depicted in Fig. 1. The $m$-th microphone signal, $m = 1, \ldots, M$, at frequency index $k$ and frame index $l$ is given by $X_m(k,l) = X_{\mathrm{d},m}(k,l) + X_{\mathrm{r},m}(k,l)$, where $X_{\mathrm{d},m}(k,l)$ denotes the direct and early reverberant speech component and $X_{\mathrm{r},m}(k,l)$ denotes the late reverberant speech component at the $m$-th microphone. In vector notation, the $M$-dimensional vec-

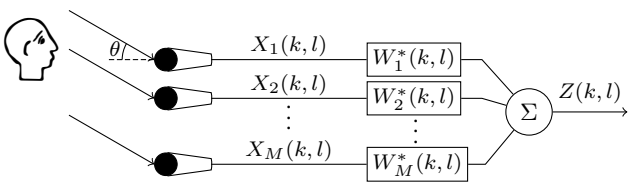


**Fig. 1**: Acoustic system configuration.

This work was supported in part by the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG), the Marie Curie Initial Training Network DREAMS (Grant no. 316969), and the joint Lower Saxony-Israeli Project ATHENA, funded by the State of Lower Saxony.

tor of the received signals $\mathbf{x}(k,l)$ can be written as

$$\mathbf{x}(k,l) = \mathbf{x}_{\mathrm{d}}(k,l) + \mathbf{x}_{\mathrm{r}}(k,l), \tag{1}$$

with $\mathbf{x}(k,l) = [X_1(k,l)\ X_2(k,l)\ \ldots\ X_M(k,l)]^T$ and $\mathbf{x}_{\mathrm{d}}(k,l)$ and $\mathbf{x}_{\mathrm{r}}(k,l)$ similarly defined. Modeling the speech source as a single point-source, the vector $\mathbf{x}_{\mathrm{d}}(k,l)$ can be expressed as

$$\mathbf{x}_{\mathrm{d}}(k,l) = S(k,l)\mathbf{d}(k), \tag{2}$$

with $S(k,l)$ the target signal (i.e., direct and early reverberant speech component) as received by a reference microphone and $\mathbf{d}(k) = [D_1(k)\ D_2(k)\ \ldots\ D_M(k)]^T$ the vector of RETFs of the target signal from the reference microphone to all microphones. The target signal is often defined as the direct speech component only, such that the vector $\mathbf{d}(k)$ can be computed based on the DOA $\theta$ of the speech source and the geometry of the microphone array [7, 8, 13–18]. The PSD matrix of $\mathbf{x}(k,l)$ is defined as

$$\mathbf{R}_{\mathbf{x}}(k,l) = \mathcal{E}\{\mathbf{x}(k,l)\mathbf{x}^H(k,l)\}, \tag{3}$$

where $\mathcal{E}$ denotes the expected value operator. Assuming that $\mathbf{x}_{\mathrm{d}}(k,l)$ and $\mathbf{x}_{\mathrm{r}}(k,l)$ are uncorrelated, the PSD matrix $\mathbf{R}_{\mathbf{x}}(k,l)$ can be written as

$$\mathbf{R}_{\mathbf{x}}(k,l) = \mathbf{R}_{\mathbf{x}_{\mathrm{d}}}(k,l) + \mathbf{R}_{\mathbf{x}_{\mathrm{r}}}(k,l), \tag{4}$$

with $\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}(k,l) = \mathcal{E}\{\mathbf{x}_{\mathrm{d}}(k,l)\mathbf{x}_{\mathrm{d}}^H(k,l)\}$ the PSD matrix of $\mathbf{x}_{\mathrm{d}}(k,l)$ and $\mathbf{R}_{\mathbf{x}_{\mathrm{r}}}(k,l) = \mathcal{E}\{\mathbf{x}_{\mathrm{r}}(k,l)\mathbf{x}_{\mathrm{r}}^H(k,l)\}$ the PSD matrix of $\mathbf{x}_{\mathrm{r}}(k,l)$. Using (2), the matrix $\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}(k,l)$ can be expressed as

$$\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}(k,l) = \Phi_{\mathrm{s}}(k,l)\mathbf{d}(k)\mathbf{d}^H(k), \tag{5}$$

where $\Phi_{\mathrm{s}}(k,l)$ is the PSD of the target signal, i.e., $\Phi_{\mathrm{s}}(k,l) = \mathcal{E}\{|S(k,l)|^2\}$. The matrix $\mathbf{R}_{\mathbf{x}_{\mathrm{r}}}(k,l)$ may be written as

$$\mathbf{R}_{\mathbf{x}_{\mathrm{r}}}(k,l) = \Phi_{\mathrm{r}}(k,l)\mathbf{\Gamma}(k), \tag{6}$$

where $\Phi_{\mathrm{r}}(k,l)$ is the PSD of the late reverberant speech component at the reference microphone and $\mathbf{\Gamma}(k)$ denotes the time-invariant spatial coherence matrix of the late reverberation normalized by $\Phi_{\mathrm{r}}(k,l)$ [7,8,13–18]. Modeling the late reverberation as an isotropic sound field, $\mathbf{\Gamma}(k)$ can be analytically computed given the geometry of the microphone array [7, 14–16]. Defining the filter coefficients vector $\mathbf{w}(k,l) = [W_1(k,l)\ W_2(k,l)\ \ldots\ W_M(k,l)]^T$, the output signal $Z(k,l)$ of the speech enhancement system is given by

$$Z(k,l) = \mathbf{w}^H(k,l)\mathbf{x}_{\mathrm{d}}(k,l) + \mathbf{w}^H(k,l)\mathbf{x}_{\mathrm{r}}(k,l). \tag{7}$$

Speech dereverberation techniques design the filter $\mathbf{w}(k,l)$ such that the output signal $Z(k,l)$ resembles the target signal $S(k,l)$. Many such techniques require an estimate of the late reverberant PSD $\Phi_{\mathrm{r}}(k,l)$, e.g., [7–9]. State-of-the-art multi-channel late reverberant PSD estimators [7, 8, 14–17] rely on knowledge of the vector $\mathbf{d}(k)$, which may be difficult to estimate accurately. In the following, a novel late reverberant PSD estimator is proposed which does not require knowledge of $\mathbf{d}(k)$. For conciseness the frequency index $k$ is omitted in the remainder of this paper.

It should be noted that for the sake of simplicity and to be able to compare the proposed estimator to the ML estimator in [8], a noise-free scenario is assumed in this paper. Nevertheless, the late reverberant PSD estimator proposed in Section 3.2 can also be used in a noisy scenario, as long as an estimate of $\mathbf{R}_{\mathbf{x}}(l)$ can be obtained by e.g., subtracting the noise component PSD matrix from the noisy signal PSD matrix.

## 3. LATE REVERBERANT POWER SPECTRAL DENSITY ESTIMATOR

In this section the ML estimator proposed in [8] is briefly reviewed and a novel eigenvalue decomposition (EVD)-based estimator is proposed.

### 3.1. Maximum likelihood estimator

In order to derive the ML estimator in [8], the spectral coefficients of the direct and late reverberant speech components are assumed to be circularly-symmetric complex Gaussian distributed. These distributions are then used to construct and maximize a likelihood function, resulting in the PSD estimates

$$\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l) = \frac{1}{M-1}\mathrm{tr}\left\{\left(\mathbf{I} - \mathbf{d}\frac{\mathbf{d}^H\mathbf{\Gamma}^{-1}}{\mathbf{d}^H\mathbf{\Gamma}^{-1}\mathbf{d}}\right)\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}\right\}, \tag{8a}$$

$$\hat{\Phi}_{\mathrm{s}}^{\mathrm{ml}}(l) = \frac{\mathbf{d}^H\mathbf{\Gamma}^{-1}}{\mathbf{d}^H\mathbf{\Gamma}^{-1}\mathbf{d}}\left[\mathbf{R}_{\mathbf{x}}(l) - \hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)\mathbf{\Gamma}\right]\frac{\mathbf{\Gamma}^{-1}\mathbf{d}}{\mathbf{d}^H\mathbf{\Gamma}^{-1}\mathbf{d}}, \tag{8b}$$

where $\mathrm{tr}\{\cdot\}$ denotes the matrix trace operator. The ML late reverberant PSD estimate in (8a) requires an estimate of the PSD matrix $\mathbf{R}_{\mathbf{x}}(l)$, normalized coherence matrix $\mathbf{\Gamma}$, and vector $\mathbf{d}$. While $\mathbf{R}_{\mathbf{x}}(l)$ can be estimated from the received signal $\mathbf{x}(l)$ and $\mathbf{\Gamma}$ can be constructed assuming a reasonable sound field model for the late reverberation, accurately estimating the vector $\mathbf{d}$ may be difficult. As is experimentally validated in [18], estimation errors in the vector $\mathbf{d}$ degrade the PSD estimation accuracy of the ML estimator in (8a), yielding as a result a degradation in the dereverberation performance of the used speech enhancement system.

### 3.2. Eigenvalue decomposition-based estimator

Aiming to remove the dependency of the PSD estimate on the vector $\mathbf{d}$, in the following we propose to estimate the late reverberant PSD using the eigenvalues of the prewhitened received signal PSD matrix $\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}$. Let us first consider the EVD of $\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}(l)\mathbf{\Gamma}^{-1}$, i.e.,

$$\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}(l)\mathbf{\Gamma}^{-1} = \mathbf{U}\mathbf{S}_{\mathbf{x}_{\mathrm{d}}}(l)\mathbf{U}^{-1}, \tag{9}$$

with $\mathbf{U}$ being an $M \times M$-dimensional matrix of eigenvectors and $\mathbf{S}_{\mathbf{x}_{\mathrm{d}}}(l)$ being the $M \times M$-dimensional diagonal matrix of eigenvalues. Since the matrix $\mathbf{R}_{\mathbf{x}_{\mathrm{d}}}(l)\mathbf{\Gamma}^{-1}$ is a rank-1 matrix, $\mathbf{S}_{\mathbf{x}_{\mathrm{d}}}(l)$ has only one non-zero eigenvalue $\sigma(l)$, i.e.,

$$\mathbf{S}_{\mathbf{x}_{\mathrm{d}}}(l) = \mathrm{diag}\{[\sigma(l)\ \ 0\ \ \ldots\ \ 0]^T\}. \tag{10}$$

Using (4) and (9), the EVD of the prewhitened received signal PSD matrix $\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}$ can be written as

$$\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1} = \mathbf{U}\mathbf{S}_{\mathbf{x}_{\mathrm{d}}}(l)\mathbf{U}^{-1} + \Phi_{\mathrm{r}}(l)\mathbf{I} \tag{11a}$$

$$= \mathbf{U}\underbrace{[\mathbf{S}_{\mathbf{x}_{\mathrm{d}}}(l) + \Phi_{\mathrm{r}}(l)\mathbf{I}]}_{\mathbf{S}_{\mathbf{x}}(l)}\mathbf{U}^{-1}, \tag{11b}$$

with $\mathbf{S}_{\mathbf{x}}(l)$ the $M \times M$-dimensional diagonal matrix of eigenvalues given by

$$\mathbf{S}_{\mathbf{x}}(l) = \mathrm{diag}\{[\sigma(l) + \Phi_{\mathrm{r}}(l)\ \ \Phi_{\mathrm{r}}(l)\ \ \ldots\ \ \Phi_{\mathrm{r}}(l)]^T\}. \tag{12}$$

Based on (11) and (12), we propose to estimate the late reverberant PSD using any of the eigenvalues of the prewhitened received signal PSD matrix $\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}$, i.e.,

$$\hat{\Phi}_{\mathrm{r}}^{\mathrm{evd}}(l) = \lambda_2\{\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}\} = \cdots = \lambda_M\{\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}\} \tag{13a}$$

$$= \frac{1}{M-1}\left(\mathrm{tr}\left\{\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}\right\} - \lambda_1\left\{\mathbf{R}_{\mathbf{x}}(l)\mathbf{\Gamma}^{-1}\right\}\right), \tag{13b}$$

where $\lambda_i\{\cdot\}$ denotes the $i$-th eigenvalue. The equality in (13b) is derived using the fact that the trace of a matrix is equal to the sum of its eigenvalues. Using $\hat{\Phi}_{\mathrm{r}}^{\mathrm{evd}}(l)$, the target signal PSD $\Phi_{\mathrm{s}}^{\mathrm{evd}}(l)$ is estimated using the decision directed approach [19].

While the ML estimate in (8a) requires knowledge of $\mathbf{R}_{\mathbf{x}}(l)$, $\mathbf{\Gamma}$, and $\mathbf{d}$, the proposed EVD-based estimate in (13) requires only knowledge of $\mathbf{R}_{\mathbf{x}}(l)$ and $\mathbf{\Gamma}$. Clearly, if $\mathbf{R}_{\mathbf{x}}(l)$ and $\mathbf{\Gamma}$ are perfectly known, the EVD-based PSD estimate in (13) is equal to the true late reverberant PSD, i.e., $\hat{\Phi}_{\mathrm{r}}^{\mathrm{evd}}(l) = \Phi_{\mathrm{r}}(l)$. In practice however, the available PSD matrix $\tilde{\mathbf{R}}_{\mathbf{x}}(l)$ and normalized coherence matrix $\tilde{\mathbf{\Gamma}}$ might differ from the true quantities $\mathbf{R}_{\mathbf{x}}(l)$ and $\mathbf{\Gamma}$. For $\tilde{\mathbf{R}}_{\mathbf{x}}(l) \neq \mathbf{R}_{\mathbf{x}}(l)$ and $\tilde{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$, the EVD of $\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}$ differs from the desired EVD in (11). Furthermore, the prewhitening of $\tilde{\mathbf{R}}_{\mathbf{x}}(l)$ using $\tilde{\mathbf{\Gamma}}^{-1}$ fails, and hence, the last $M-1$ eigenvalues of $\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}$ are not equal, i.e.,

$$\lambda_2\{\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}\} \neq \lambda_3\{\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}\} \neq \ldots \neq \lambda_M\{\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}\}. \tag{14}$$

As a result, different late reverberant PSD estimates are obtained depending on the eigenvalue used for the estimation. In the remainder of this work, $\hat{\Phi}_{\mathrm{r},i}^{\mathrm{evd}}(l)$ will be used to denote the late reverberant PSD estimate when using the $i$-th eigenvalue of $\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}$, i.e.,

$$\hat{\Phi}_{\mathrm{r},i}^{\mathrm{evd}}(l) = \lambda_i\left\{\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}\right\}, \quad i = 2, \ldots, M. \tag{15}$$

Furthermore, $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$ will be used to denote the late reverberant PSD estimate when using the trace and the first eigenvalue of $\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}$, i.e.,

$$\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l) = \frac{1}{M-1}\left(\mathrm{tr}\left\{\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}\right\} - \lambda_1\left\{\tilde{\mathbf{R}}_{\mathbf{x}}(l)\tilde{\mathbf{\Gamma}}^{-1}\right\}\right). \tag{16}$$

In Section 4 the performance of the PSD estimate in (15) with $i = 2$ as well as of the PSD estimate in (16) is investigated.

## 4. EXPERIMENTAL RESULTS

In this section the dereverberation performance of a MWF using the considered ML and EVD-based PSD estimators is compared when the DOA is perfectly estimated as well as in the presence of DOA estimation errors. As in e.g., [7, 8], the MWF is implemented as an MVDR beamformer $\mathbf{w}_{\mathrm{MVDR}}$ followed by a single-channel Wiener postfilter $G(l)$ applied to the MVDR output, i.e.,

$$\mathbf{w}_{\mathrm{MWF}}(l) = \underbrace{\frac{\tilde{\mathbf{\Gamma}}^{-1}\mathbf{d}}{\mathbf{d}^H\tilde{\mathbf{\Gamma}}^{-1}\mathbf{d}}}_{\mathbf{w}_{\mathrm{MVDR}}} \underbrace{\frac{\Phi_{\mathrm{s_o}}(l)}{\Phi_{\mathrm{s_o}}(l) + \Phi_{\mathrm{r_o}}(l)}}_{G(l)}, \tag{17}$$

with $\Phi_{\mathrm{s_o}}(l)$ and $\Phi_{\mathrm{r_o}}(l)$ the PSDs of the target signal and late reverberant speech component at the output of the MVDR beamformer given by

$$\Phi_{\mathrm{s_o}}(l) = \Phi_{\mathrm{s}}(l), \quad \Phi_{\mathrm{r_o}}(l) = \frac{\Phi_{\mathrm{r}}(l)}{\mathbf{d}^H\tilde{\mathbf{\Gamma}}^{-1}\mathbf{d}}. \tag{18}$$

While the PSD of the target signal is not changed by the MVDR beamformer, the PSD of the late reverberant speech component needs to be corrected by the beamformer suppression factor. The MVDR beamformer $\mathbf{w}_{\mathrm{MVDR}}$ in (17) is time-invariant and depends on the available normalized coherence matrix $\tilde{\mathbf{\Gamma}}$ and vector $\mathbf{d}$. On the other hand, the single-channel Wiener filter $G(l)$ depends on the time-varying PSDs $\Phi_{\mathrm{s}}(l)$ and $\Phi_{\mathrm{r}}(l)$, cf. (17) and (18), which are estimated using the considered ML and EVD-based estimators. It should be noted that independently of the PSD estimator used, the MWF implemented according to (17) is sensitive to errors in the vector $\mathbf{d}$ due to the sensitivity of the MVDR beamformer to errors in $\mathbf{d}$. However, as shown in Section 4.3, a significantly higher sensitivity of the MWF is observed when also the used PSD estimator is affected by errors in the vector $\mathbf{d}$.

### 4.1. Setup

We consider two multi-channel acoustic systems with $M \in \{2, 3, 4\}$ microphones and a single speech source located at an angle $\theta = 45^\circ$ and at a distance of 2 m from the microphone array. The first acoustic system consists of a linear microphone array with an inter-sensor distance of 8 cm placed in a room with reverberation time $T_{60} \approx 610$ ms [20]. The second acoustic system consists of a circular microphone array with a radius of 10 cm placed in a room with reverberation time $T_{60} \approx 730$ ms [21]. The sampling frequency is $f_s = 16$ kHz and the received reverberant signals are generated by convolving clean speech signals from the HINT database [22] with the measured RIRs.

The signals are processed using a weighted overlap-add framework with a frame size of 1024 samples and an overlap of 75% between successive frames. The first microphone is arbitrarily selected as the reference microphone. In order to implement the MWF, the vector $\mathbf{d}$, the normalized coherence matrix $\tilde{\mathbf{\Gamma}}$, and the received signal PSD matrix $\tilde{\mathbf{R}}_{\mathbf{x}}(l)$ are required. As in e.g., [8, 13, 18], the vector $\mathbf{d}$ is computed from the respective RIRs truncated to the part containing only the direct path response. As in e.g., [7, 14], the normalized coherence matrix $\tilde{\mathbf{\Gamma}}$ is computed assuming a spherically isotropic sound field. In none of the considered acoustic systems the late reverberation is truly isotropic, resulting in a mismatch between the available coherence matrix $\tilde{\mathbf{\Gamma}}$ and the true coherence matrix $\mathbf{\Gamma}$. The received signal PSD matrix $\tilde{\mathbf{R}}_{\mathbf{x}}(l)$ is estimated from $\mathbf{x}(l)$ using recursive averaging with a time constant of 40 ms. The minimum gain of the single-channel Wiener postfilter is set to $-20$ dB.

The performance is evaluated in terms of the improvement in frequency-weighted segmental signal-to-noise-ratio ($\Delta$fwSSNR) [23] and the improvement in cepstral distance ($\Delta$CD) [24] between the output speech signal and the reference microphone signal. The fwSSNR and CD measures are intrusive measures comparing the output signal to a reference signal. The reference signal used in this paper is the anechoic speech signal. It should be noted that a positive $\Delta$fwSSNR and a negative $\Delta$CD indicate a performance improvement.

The performance of the MVDR beamformer in (17) and of the MWF implemented according to (17) using the ML and EVD-based PSD estimates is investigated for

i) both acoustic systems with different number of microphones $M \in \{2, 3, 4\}$ and assuming *the DOA is perfectly estimated*, i.e., the vector $\mathbf{d}$ is computed from the truncated RIRs corresponding to the true DOA $\theta = 45^\circ$ (Section 4.2),

ii) the first acoustic system with $M = 4$ microphones and assuming *DOA estimation errors*, i.e., the vector $\mathbf{d}$ is computed from the truncated RIRs corresponding to several erroneous DOAs $\hat{\theta} \in \{-90, -75, \ldots, 90\}$. In other words, the actual source position is always at $\theta = 45^\circ$, but the vector $\mathbf{d}$ is computed using different DOAs $\hat{\theta}$ (Section 4.3).

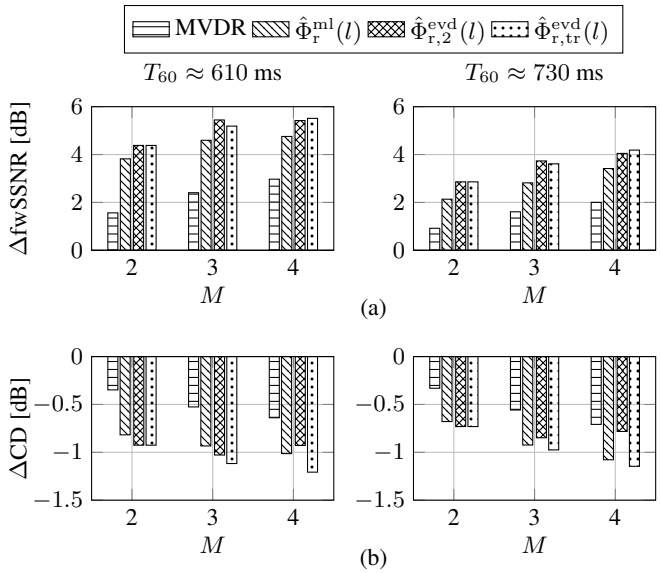


**Fig. 2**: Performance of the MVDR beamformer and the MWF using $\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)$, $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$, and $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$ when the DOA is perfectly estimated: (a) ΔfwSSNR and (b) ΔCD.

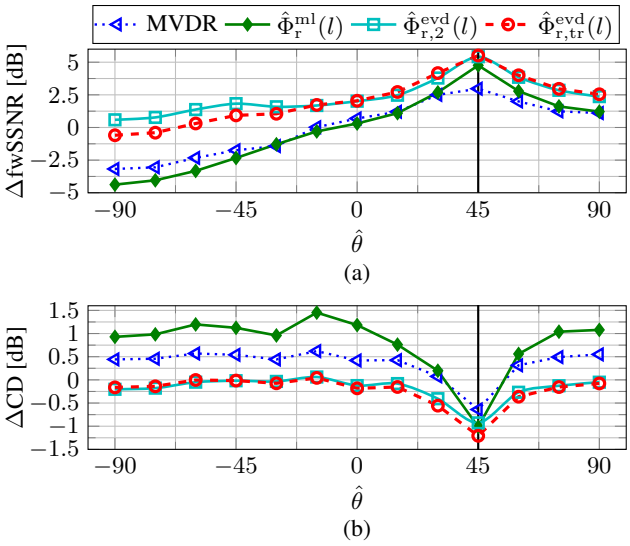


**Fig. 3**: Performance of the MVDR beamformer and the MWF using $\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)$, $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$, and $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$ in the presence of DOA estimation errors: (a) ΔfwSSNR and (b) ΔCD ($T_{60} \approx 610$ ms, $M = 4$).

### 4.2. Performance when the DOA is perfectly estimated

Fig. 2 depicts the ΔfwSSNR and ΔCD obtained for the MVDR beamformer and the MWF using late reverberant PSD estimates $\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)$, $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$, and $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$. As expected, it can be observed that the MWF using any of the considered PSD estimates improves the performance in comparison to the MVDR beamformer. In addition, it can be observed that the performance of the MVDR beamformer and of the MWF using any of the considered PSD estimates increases with increasing number of microphones (except for the ΔCD obtained when using the MWF with $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$ being slightly worse for $M = 4$ than for $M = 3$). For both acoustic systems and all considered configurations, the proposed EVD-based PSD estimate $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$ typically yields the best performance, always outperforming the ML estimate $\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)$. Furthermore, the EVD-based estimate $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$ also outperforms the ML estimate $\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)$ in terms of the ΔfwSSNR, while a lower performance is obtained in terms of the ΔCD for $M = 4$. It should be noted that for $M = 2$, $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l) = \hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$, cf. (15) and (16), and hence the performance when using these EVD-based PSD estimates is the same.

In summary, it can be said that among the proposed EVD-based PSD estimates $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$ and $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$, using $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$ in a MWF generally yields a better performance. In addition, the proposed EVD-based PSD estimate $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$ yields a better performance than the ML estimate $\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)$.

### 4.3. Performance in the presence of DOA estimation errors

In order to investigate the performance of the MVDR beamformer and of the MWF in the presence of DOA estimation errors, Fig. 3 depicts the ΔfwSSNR and ΔCD obtained when the vector $\mathbf{d}$ is computed from the truncated RIRs corresponding to several erroneous DOAs $\hat{\theta}$. For completeness, the performance when the vector $\mathbf{d}$ is computed from the truncated RIRs corresponding to the true DOA, i.e., $\hat{\theta} = \theta = 45°$, is also depicted. The presented ΔfwSSNR and ΔCD show that as expected, DOA estimation errors yield a performance degradation for the MVDR beamformer. Furthermore, it can be observed that since the MWF is equivalent to the MVDR beamformer followed by a postfilter, cf. (17), DOA estimation errors yield a performance degradation also for the MWF using any of the considered PSD estimates. Since the ML estimate additionally relies on knowledge of the DOA, using this PSD estimate for the MWF results in a significantly faster and larger performance degradation than using the proposed EVD-based PSD estimates. Since the proposed EVD-based PSD estimates are independent of the DOA, the performance degradation of the MWF when using these PSD estimates occurs merely due to the sensitivity of the MVDR beamformer to DOA estimation errors. The performance when using proposed EVD-based PSD estimates $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$, and $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$ is very similar, with $\hat{\Phi}_{\mathrm{r,tr}}^{\mathrm{evd}}(l)$ yielding a slightly better performance for small DOA estimation errors and $\hat{\Phi}_{\mathrm{r},2}^{\mathrm{evd}}(l)$ yielding a slightly better performance for large DOA estimation errors.

In summary, it can be said that since the proposed EVD-based PSD estimates do not require knowledge of the DOA, using any of these PSD estimates for the MWF results in significantly better performance than using the ML estimate in the presence of DOA estimation errors.

## 5. CONCLUSION

In this paper a novel multi-channel late reverberant PSD estimator has been proposed which does not require an estimate of the DOA of the speech source or of the RETFs of the target signal. Modeling reverberation as an isotropic sound field, it has been proposed to estimate the late reverberant PSD based on the eigenvalues of the prewhitened received signal PSD matrix. Experimental results have shown that using the proposed EVD-based PSD estimator in a MWF for speech dereverberation yields a better performance than a recently proposed ML estimator, both when the DOA is perfectly estimated as well as in the presence of DOA estimation errors.

## 6. REFERENCES


[1] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, July 2006.

[2] S. Goetze, E. Albertin, J. Rennies, E. A. P. Habets, and K.-D. Kammeyer, "Speech quality assessment for listening-room compensation," in *Proc. AES International Conference on Sound Quality Evaluation*, Pitea, Sweden, June 2010, pp. 11–20.

[3] A. Warzybok, I. Kodrasi, J. O. Jungmann, E. A. P. Habets, T. Gerkmann, A. Mertins, S. Doclo, B. Kollmeier, and S. Goetze, "Subjective speech quality and speech intelligibility evaluation of single-channel dereverberation algorithm," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sept. 2014, pp. 333–337.

[4] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, Nov. 2012.

[5] F. Xiong, B. T. Meyer, N. Moritz, R. Rehr, J. Anemüller, T. Gerkmann, S. Doclo, and S. Goetze, "Front-end technologies for robust ASR in reverberant environments–spectral enhancement-based dereverberation and auditory modulation filterbank features," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, pp. 1–18, Aug. 2015.

[6] P. A. Naylor and N. D. Gaubitch, Eds., *Speech dereverberation*, Springer, London, UK, 2010.

[7] S. Braun and E. A. P. Habets, "Dereverberation in noisy environments using reference signals and a maximum likelihood estimator," in *Proc. European Signal Processing Conference*, Marrakech, Morocco, Sept. 2013.

[8] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," in *Proc. European Signal Processing Conference*, Lisbon, Portugal, Sept. 2014.

[9] O. Schwartz, S. Gannot, and E. A. P. Habets, "Multi-microphone speech dereverberation and noise reduction using relative early transfer functions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 2, pp. 240–251, Feb. 2015.

[10] J. Jensen and M. S. Pedersen, "Analysis of beamformer directed single-channel noise reduction system for hearing aid applications," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Brisbane, Australia, Apr. 2015, pp. 5728–5732.

[11] O. Thiergart, G. Del Galdo, and E. A. P. Habets, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation," *Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2337–2346, Oct. 2012.

[12] A. Schwarz and W. Kellermann, "Coherent-to-diffuse power ratio estimation for dereverberation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 6, pp. 1006–1018, June 2015.

[13] A. Kuklasiński, S. Doclo, T. Gerkmann, S. H. Jensen, and J. Jensen, "Multi-channel PSD estimators for speech dereverberation - a theoretical and experimental comparison," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Brisbane, Australia, Apr. 2015, pp. 91–95.

[14] S. Braun and E. A. P. Habets, "A multichannel diffuse power estimator for dereverberation in the presence of multiple sources," *EURASIP Journal on Applied Signal Processing*, vol. 2015, no. 1, pp. 1–14, Dec. 2015.

[15] O. Schwartz, S. Braun, S. Gannot, and E. A. P. Habets, "Maximum likelihood estimation of the late reverberant power spectral density in noisy environments," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, Oct. 2015.

[16] O. Schwartz, S. Gannot, and E. A. P. Habets, "Joint maximum likelihood estimation of late reverberant and speech power spectral density in noisy environments," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016, pp. 151–155.

[17] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum Likelihood PSD estimation for speech enhancement in reverberation and noise," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1595–1608, Sept. 2016.

[18] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Multichannel Wiener filter for speech dereverberation in hearing aids - sensitivity to DoA errors," in *Proc. AES 60th Conference on Dereverberation and Reverberation of Audio, Music, and Speech*, Leuven, Belgium, Feb. 2016.

[19] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[20] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sept. 2014, pp. 313–317.

[21] K. Kinoshita, M. Delcroix, S. Gannot, E. A. P. Habets, R. Haeb-Umbach, W. Kellermann, V. Leutnant, R. Maas, T. Nakatani, B. Raj, A. Sehr, and T. Yoshioka, "A summary of the REVERB challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–19, Jan. 2016.

[22] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1085–1099, Feb. 1994.

[23] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[24] S. Quackenbush, T. Barnwell, and M. Clements, *Objective measures of speech quality*, Prentice-Hall, New Jersey, USA, 1988.