

RETF-Independent Multi-Channel Late Reverberant Power Spectral Density Estimator

Ina Kodrasi, *Member, IEEE*, Simon Doclo, *Senior Member, IEEE*

Abstract—

Index Terms—

I. INTRODUCTION

THIS

II. CONFIGURATION AND NOTATION

Consider a reverberant and noisy acoustic system with a single speech source and $M \geq 2$ microphones as depicted in Fig. 1. The m -th microphone signal $Y_m(k, l)$ at frequency index k and frame index l is given by

$$Y_m(k, l) = X_m(k, l) + V_m(k, l), \quad m = 1, 2, \dots, M, \quad (1)$$

with $X_m(k, l)$ the reverberant speech component and $V_m(k, l)$ the additive noise component. Denoting the direct and early reverberant speech component by $X_{d,m}(k, l)$ and the late reverberant speech component by $X_{r,m}(k, l)$, the m -th microphone signal in (1) can be expressed as

$$Y_m(k, l) = \underbrace{X_{d,m}(k, l) + X_{r,m}(k, l)}_{X_m(k, l)} + V_m(k, l). \quad (2)$$

In vector notation, the M -dimensional vector of the microphone signals $\mathbf{y}(k, l)$ can be written as

$$\mathbf{y}(k, l) = \underbrace{\mathbf{x}_d(k, l) + \mathbf{x}_r(k, l)}_{\mathbf{x}(k, l)} + \mathbf{v}(k, l), \quad (3)$$

The authors are with the Signal Processing Group, Department of Medical Physics and Acoustics, and Cluster of Excellence Hearing4All, University of Oldenburg, Germany (email: ina.kodrasi@uni-oldenburg.de, simon.doclo@uni-oldenburg.de).

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. Manuscript received xx; revised xx.

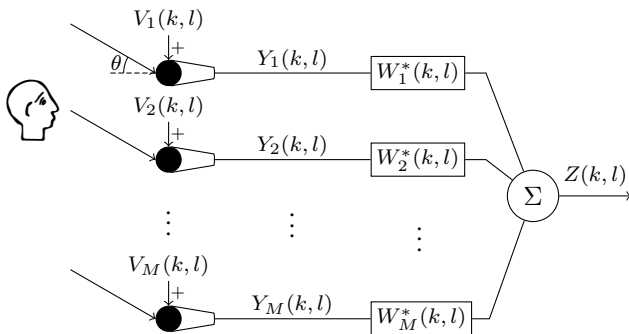


Fig. 1: Acoustic system configuration.

with $\mathbf{y}(k, l) = [Y_1(k, l) \ Y_2(k, l) \ \dots \ Y_M(k, l)]^T$ and $\mathbf{x}(k, l)$, $\mathbf{x}_d(k, l)$, $\mathbf{x}_r(k, l)$, and $\mathbf{v}(k, l)$ similarly defined.

Modelling the speech source as a single point-source, the direct and early reverberant speech component vector $\mathbf{x}_d(k, l)$ is equal to

$$\mathbf{x}_d(k, l) = S(k, l)\mathbf{d}(k), \quad (4)$$

with $S(k, l)$ the target signal (i.e., direct and early reverberant speech component) as received by a reference microphone and

$$\mathbf{d}(k) = [D_1(k) \ D_2(k) \ \dots \ D_M(k)]^T, \quad (5)$$

the vector of RETFs of the target signal from the reference microphone to all microphones. For simplicity, 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 θ of the speech source and the geometry of the microphone array [1]–[8].

The PSD matrix of the microphone signals $\mathbf{y}(k, l)$ is defined as

$$\Phi_{\mathbf{y}}(k, l) = \mathcal{E}\{\mathbf{y}(k, l)\mathbf{y}^H(k, l)\}, \quad (6)$$

with \mathcal{E} the expected value operator. Assuming that the speech and noise components are uncorrelated, the PSD matrix $\Phi_{\mathbf{y}}(k, l)$ can be written as

$$\Phi_{\mathbf{y}}(k, l) = \Phi_{\mathbf{x}}(k, l) + \Phi_{\mathbf{v}}(k, l), \quad (7)$$

with $\Phi_{\mathbf{x}}(k, l) = \mathcal{E}\{\mathbf{x}(k, l)\mathbf{x}^H(k, l)\}$ the PSD matrix of the reverberant speech component and $\Phi_{\mathbf{v}}(k, l) = \mathcal{E}\{\mathbf{v}(k, l)\mathbf{v}^H(k, l)\}$ the PSD matrix of the noise component. Furthermore, assuming that $\mathbf{x}_d(k, l)$ and $\mathbf{x}_r(k, l)$ are uncorrelated, $\Phi_{\mathbf{x}}(k, l)$ can be written as

$$\Phi_{\mathbf{x}}(k, l) = \Phi_{\mathbf{x}_d}(k, l) + \Phi_{\mathbf{x}_r}(k, l), \quad (8)$$

with $\Phi_{\mathbf{x}_d}(k, l) = \mathcal{E}\{\mathbf{x}_d(k, l)\mathbf{x}_d^H(k, l)\}$ the PSD matrix of the direct and early reverberant speech component and $\Phi_{\mathbf{x}_r}(k, l) = \mathcal{E}\{\mathbf{x}_r(k, l)\mathbf{x}_r^H(k, l)\}$ the PSD matrix of the late reverberant speech component. Using (4), the matrix $\Phi_{\mathbf{x}_d}(k, l)$ is equal to

$$\Phi_{\mathbf{x}_d}(k, l) = \Phi_s(k, l)\mathbf{d}(k)\mathbf{d}^H(k), \quad (9)$$

with $\Phi_s(k, l)$ the PSD of the target signal, i.e., $\Phi_s(k, l) = \mathcal{E}\{|S(k, l)|^2\}$. The matrix $\Phi_{\mathbf{x}_r}(k, l)$ may be written as

$$\Phi_{\mathbf{x}_r}(k, l) = \Phi_r(k, l)\Gamma(k), \quad (10)$$

with $\Phi_r(k, l)$ the PSD of the late reverberant speech component at the reference microphone and $\Gamma(k)$ the time-invariant spatial coherence matrix of the late reverberation normalized by $\Phi_r(k, l)$ [1]–[8]. Modelling the late reverberation as an

isotropic sound field, $\Gamma(k)$ can be analytically computed given the geometry of the microphone array [1], [4]–[6].

Defining the filter coefficients vector $\mathbf{w}(k, l) = [W_1(k, l) \ W_2(k, l) \ \dots \ W_M(k, l)]^T$, the output signal $Z(k, l)$ of the speech enhancement system is given by

$$Z(k, l) = \mathbf{w}^H(k, l)\mathbf{x}_d(k, l) + \mathbf{w}^H(k, l)\mathbf{x}_r(k, l) + \mathbf{w}^H(k, l)\mathbf{v}(k, l). \quad (11)$$

Dereverberation and noise reduction techniques aim at designing speech enhancement filters $\mathbf{w}(k, l)$ 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 $\Phi_r(k, l)$, e.g., [1], [2], [9]. State-of-the-art multi-channel late reverberant PSD estimators [1], [2], [4]–[7] rely on knowledge of the RETF vector $\mathbf{d}(k)$, which may be difficult to estimate accurately, particularly in highly reverberant and noisy environments. As is experimentally validated in e.g. [8], [10], an erroneously estimated $\mathbf{d}(k)$ degrades the PSD estimation accuracy, yielding as a result a degradation in the performance of the speech enhancement system. In this paper a novel multi-channel late reverberant PSD estimator is proposed which does not require knowledge of the RETF vector $\mathbf{d}(k)$. For conciseness, the frequency index k and the frame index l are omitted in the following unless explicitly required.

III. EVD-BASED LATE REVERBERANT POWER SPECTRAL DENSITY ESTIMATOR

In this section the ML estimator proposed in [2] is briefly reviewed and the novel EVD-based estimator is described. In addition, it is analytically shown that when the RETF vector \mathbf{d} and the normalized coherence matrix Γ are perfectly known, the ML and EVD-based estimators are equivalent. For the sake of simplicity and to be able to compare the proposed EVD-based estimator to the ML estimator, a noise-free scenario is assumed, i.e., $\mathbf{y} = \mathbf{x}$. Nevertheless, the EVD-based PSD estimator proposed in Section III-B can also be used in a noisy scenario (cf. Section V-E), as long as an estimate of the reverberant speech component PSD matrix Φ_x can be obtained (e.g., by subtracting the noise component PSD matrix Φ_v from the noisy signal PSD matrix Φ_y using a voice activity detector).

A. ML estimator

In order to derive the ML estimator in [2], the spectral coefficients of the direct and late reverberant speech components are assumed to be circularly-symmetric complex Gaussian distributed. These distributions are used to construct and maximize a likelihood function, resulting in the PSD estimates

$$\hat{\Phi}_r^{\text{ml}} = \frac{1}{M-1} \text{tr} \left\{ \left(\mathbf{I} - \mathbf{d} \frac{\mathbf{d}^H \Gamma^{-1}}{\mathbf{d}^H \Gamma^{-1} \mathbf{d}} \right) \Phi_x \Gamma^{-1} \right\}, \quad (12a)$$

$$\hat{\Phi}_s^{\text{ml}} = \frac{\mathbf{d}^H \Gamma^{-1}}{\mathbf{d}^H \Gamma^{-1} \mathbf{d}} \left(\Phi_x - \hat{\Phi}_r^{\text{ml}} \Gamma \right) \frac{\Gamma^{-1} \mathbf{d}}{\mathbf{d}^H \Gamma^{-1} \mathbf{d}}, \quad (12b)$$

where $\text{tr}\{\cdot\}$ denotes the matrix trace operator and \mathbf{I} denotes the $M \times M$ -dimensional identity matrix. The ML late reverberant PSD estimate in (12a) requires an estimate of the PSD matrix

Φ_x , normalized coherence matrix Γ , and vector \mathbf{d} . While Φ_x can be estimated from the reverberant speech component \mathbf{x} (or e.g., by subtracting Φ_v from Φ_y in a noisy scenario) and Γ can be constructed assuming a reasonable sound field model for the late reverberation, accurately estimating the vector \mathbf{d} may be difficult, particularly in highly reverberant and noisy environments. As is experimentally validated in [8], an erroneously estimated \mathbf{d} degrades the accuracy of the ML PSD estimate in (12a), yielding as a result a degradation in the performance of the speech enhancement system.

B. EVD-based estimator

Aiming to remove the dependency of the PSD estimate on the RETF vector \mathbf{d} , we propose to estimate the late reverberant PSD using the eigenvalues of the prewhitened reverberant speech component PSD matrix $\Phi_x \Gamma^{-1}$.

Let us first consider the EVD of $\Phi_{x_d} \Gamma^{-1}$, i.e.,

$$\Phi_{x_d} \Gamma^{-1} = \mathbf{U} \mathbf{S}_{x_d} \mathbf{U}^{-1}, \quad (13)$$

with \mathbf{U} being an $M \times M$ -dimensional matrix of eigenvectors and \mathbf{S}_{x_d} being the $M \times M$ -dimensional diagonal matrix of eigenvalues. Since the matrix $\Phi_{x_d} \Gamma^{-1}$ is a rank-1 matrix, \mathbf{S}_{x_d} has only one non-zero eigenvalue σ , i.e.,

$$\mathbf{S}_{x_d} = \text{diag}\{\sigma \ 0 \ \dots \ 0\}^T. \quad (14)$$

Based on (8) and (13), the EVD of the prewhitened reverberant speech component PSD matrix $\Phi_x \Gamma^{-1}$ can be written as

$$\Phi_x \Gamma^{-1} = \mathbf{U} \mathbf{S}_{x_d} \mathbf{U}^{-1} + \Phi_r \mathbf{I} \quad (15a)$$

$$= \mathbf{U} \underbrace{[\mathbf{S}_{x_d} + \Phi_r \mathbf{I}]}_{\mathbf{S}_x} \mathbf{U}^{-1}, \quad (15b)$$

with \mathbf{S}_x the $M \times M$ -dimensional diagonal matrix of eigenvalues given by

$$\mathbf{S}_x = \text{diag}\{[\sigma + \Phi_r \ \Phi_r \ \dots \ \Phi_r]^T\}. \quad (16)$$

Observing (15) and (16), we propose to estimate the late reverberant PSD using any of the eigenvalues of the matrix $\Phi_x \Gamma^{-1}$, i.e.,

$$\hat{\Phi}_r^{\text{evd}} = \lambda_2\{\Phi_x \Gamma^{-1}\} = \dots = \lambda_M\{\Phi_x \Gamma^{-1}\} \quad (17a)$$

$$= \frac{1}{M-1} (\text{tr}\{\Phi_x \Gamma^{-1}\} - \lambda_1\{\Phi_x \Gamma^{-1}\}), \quad (17b)$$

where $\lambda_i\{\cdot\}$ denotes the i -th eigenvalue. The equality in (17b) is derived using the fact that the trace of a matrix is equal to the sum of its eigenvalues. Using $\hat{\Phi}_r^{\text{evd}}$, the target signal PSD $\hat{\Phi}_s^{\text{evd}}$ can be estimated e.g., using the decision directed approach [11].

While the ML estimate in (12a) requires knowledge of Φ_x , Γ , and \mathbf{d} , the proposed EVD-based estimate in (17) requires only knowledge of Φ_x and Γ . If Γ is perfectly known, the EVD-based PSD estimate in (17) is clearly equal to the true late reverberant PSD, i.e., $\hat{\Phi}_r^{\text{evd}} = \Phi_r$.

C. Equivalence of the ML and EVD-based PSD estimators for perfect knowledge of the RETF vector and coherence matrix

In this section we analytically show that when the true RETF vector \mathbf{d} and coherence matrix $\mathbf{\Gamma}$ are available, the ML estimate in (12a) and the EVD-based estimate in (17) are equivalent and equal to the true late reverberant PSD.

The ML PSD estimate in (12a) can be written as

$$\hat{\Phi}_r^{\text{ml}} = \frac{1}{M-1} \left(\text{tr} \{ \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \} - \text{tr} \left\{ \frac{\mathbf{d} \mathbf{d}^H \mathbf{\Gamma}^{-1}}{\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d}} \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \right\} \right). \quad (18)$$

Using (8), the second term in (18) is equal to

$$\text{tr} \left\{ \frac{\mathbf{d} \mathbf{d}^H \mathbf{\Gamma}^{-1}}{\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d}} \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \right\} \quad (19)$$

$$= \text{tr} \left\{ \frac{\mathbf{d} \mathbf{d}^H \mathbf{\Gamma}^{-1}}{\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d}} \Phi_s \mathbf{d} \mathbf{d}^H \mathbf{\Gamma}^{-1} \right\} + \text{tr} \left\{ \frac{\mathbf{d} \mathbf{d}^H \mathbf{\Gamma}^{-1}}{\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d}} \Phi_r \right\} \quad (20)$$

$$= \text{tr} \left\{ \frac{(\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d})^2}{\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d}} \Phi_s \right\} + \text{tr} \left\{ \frac{\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d}}{\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d}} \Phi_r \right\} \quad (21)$$

$$= \text{tr} \{ \mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d} \Phi_s \} + \Phi_r \quad (22)$$

$$= \text{tr} \{ \mathbf{\Phi}_{x_d} \mathbf{\Gamma}^{-1} \} + \Phi_r. \quad (23)$$

The equalities in (21)-(23) are derived using the fact that the trace is invariant under cyclic permutations. Since the trace of a matrix is equal to the sum of its eigenvalues and since the matrix $\mathbf{\Phi}_{x_d} \mathbf{\Gamma}^{-1}$ has only one non-zero eigenvalue σ , cf. (14), (23) is equal to

$$\text{tr} \{ \mathbf{\Phi}_{x_d} \mathbf{\Gamma}^{-1} \} + \Phi_r = \sigma + \Phi_r = \lambda_1 \{ \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \}. \quad (24)$$

Based on (24), the ML PSD estimate in (18) can be written as

$$\hat{\Phi}_r^{\text{ml}} = \frac{1}{M-1} (\text{tr} \{ \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \} - \lambda_1 \{ \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \}) \quad (25)$$

$$= \hat{\Phi}_r^{\text{evd}} = \Phi_r. \quad (26)$$

Hence, when the RETF vector \mathbf{d} and normalized coherence matrix $\mathbf{\Gamma}$ are perfectly known, the ML and EVD-based PSD estimates are equivalent and equal to the true late reverberant PSD Φ_r . These theoretical results are experimentally validated in Section V-B.

IV. ANALYTICAL COMPARISON OF THE ML AND EVD-BASED ESTIMATORS IN THE PRESENCE OF RETF VECTOR AND COHERENCE MATRIX MISMATCHES

In Section III-C it was shown that when the RETF vector \mathbf{d} and normalized coherence matrix $\mathbf{\Gamma}$ are perfectly known, the ML and EVD-based PSD estimates are equivalent. In practice however, the available RETF vector $\tilde{\mathbf{d}}$ and coherence matrix $\tilde{\mathbf{\Gamma}}$ might differ from the true \mathbf{d} and $\mathbf{\Gamma}$, i.e.,

$$\tilde{\mathbf{d}} = \mathbf{d} + \Delta \mathbf{d}, \quad \tilde{\mathbf{\Gamma}} = \mathbf{\Gamma} + \Delta \mathbf{\Gamma}, \quad (27)$$

with $\Delta \mathbf{d} \neq \mathbf{0}$ and $\Delta \mathbf{\Gamma} \neq \mathbf{0}$ the mismatch between the true and the available quantities. In the presence of RETF vector or coherence matrix mismatches, the ML and EVD-based PSD estimates are not equivalent and the derivations provided in Section III-C do not hold. In the following, the effect of RETF vector and coherence matrix mismatches on both PSD estimates is theoretically analyzed.

A. RETF vector mismatches

In this section it is analytically shown that the ML estimator overestimates the late reverberant PSD in the presence of RETF vector mismatches. Clearly, RETF vector mismatches do not affect the proposed EVD-based estimator, cf. (17).

Using $\tilde{\mathbf{d}}$ instead of \mathbf{d} in (12a), the ML late reverberant PSD estimate is given by

$$\hat{\Phi}_r^{\text{ml}} = \frac{1}{M-1} \left(\text{tr} \{ \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \} - \text{tr} \left\{ \frac{\tilde{\mathbf{d}} \tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1}}{\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}} \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \right\} \right), \quad (28)$$

Similarly as before, the second term in (28) can be written as

$$\text{tr} \left\{ \frac{\tilde{\mathbf{d}} \tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1}}{\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}} \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \right\} \quad (29)$$

$$= \text{tr} \left\{ \frac{\tilde{\mathbf{d}} \tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1}}{\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}} \Phi_s \mathbf{d} \mathbf{d}^H \mathbf{\Gamma}^{-1} \right\} + \text{tr} \left\{ \frac{\tilde{\mathbf{d}} \tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1}}{\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}} \Phi_r \right\} \quad (30)$$

$$= \frac{(\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \mathbf{d})^2}{\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}} \Phi_s + \Phi_r, \quad (31)$$

such that the ML estimate in (28) is equal to

$$\hat{\Phi}_r^{\text{ml}} = \frac{1}{M-1} \left(\text{tr} \{ \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \} - \frac{(\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \mathbf{d})^2}{\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}} \Phi_s - \Phi_r \right). \quad (32)$$

Since the true late reverberant PSD can be written as (cf. (18), (22), (26))

$$\Phi_r = \frac{1}{M-1} (\text{tr} \{ \mathbf{\Phi}_x \mathbf{\Gamma}^{-1} \} - \mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d} \Phi_s - \Phi_r), \quad (33)$$

the difference between the ML estimate in (32) and the true PSD in (33) is equal to

$$\Delta \hat{\Phi}_r^{\text{ml}} = \hat{\Phi}_r^{\text{ml}} - \Phi_r \quad (34)$$

$$= \frac{\Phi_s}{M-1} \frac{(\mathbf{d}^H \mathbf{\Gamma}^{-1} \mathbf{d})(\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}) - (\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \mathbf{d})^2}{\tilde{\mathbf{d}}^H \mathbf{\Gamma}^{-1} \tilde{\mathbf{d}}}. \quad (35)$$

In the following it is shown that $\Delta \hat{\Phi}_r^{\text{ml}} > 0$, i.e., the ML estimator always yields an overestimation of the late reverberant PSD in the presence of RETF vector mismatches.

Since $\mathbf{\Gamma}^{-1}$ is the inverse of a coherence matrix, it is symmetric and positive definite and can be decomposed using the Cholesky decomposition as

$$\mathbf{\Gamma}^{-1} = \mathbf{L} \mathbf{L}^H, \quad (36)$$

with \mathbf{L} an $M \times M$ lower triangular matrix. Using (36), $\Delta \hat{\Phi}_r^{\text{ml}}$ in (35) can be written as

$$\Delta \hat{\Phi}_r^{\text{ml}} = \frac{\Phi_s}{M-1} \frac{(\mathbf{d}^H \mathbf{L} \mathbf{L}^H \mathbf{d})(\tilde{\mathbf{d}}^H \mathbf{L} \mathbf{L}^H \tilde{\mathbf{d}}) - (\tilde{\mathbf{d}}^H \mathbf{L} \mathbf{L}^H \mathbf{d})^2}{\tilde{\mathbf{d}}^H \mathbf{L} \mathbf{L}^H \tilde{\mathbf{d}}} \quad (37)$$

$$= \frac{\Phi_s}{M-1} \frac{(\mathbf{a}^H \mathbf{a})(\mathbf{b}^H \mathbf{b}) - (\mathbf{b}^H \mathbf{a})^2}{\mathbf{b}^H \mathbf{b}}, \quad (38)$$

with the vectors $\mathbf{a} = \mathbf{L}^H \mathbf{d}$ and $\mathbf{b} = \mathbf{L}^H \tilde{\mathbf{d}}$ introduced in order to simplify the notation. Using the Cauchy-Schwarz inequality, it can be said that

$$(\mathbf{a}^H \mathbf{a})(\mathbf{b}^H \mathbf{b}) - (\mathbf{b}^H \mathbf{a})^2 > 0. \quad (39)$$

Based on (38), the difference between the ML estimate and the true late reverberant PSD in (38) is always greater than 0, i.e.,

$$\Delta\hat{\Phi}_r^{\text{ml}} > 0. \quad (40)$$

Hence, in the presence of RETF vector mismatches the ML estimator overestimates the late reverberant PSD. This theoretical result is experimentally validated in Section V-B.

B. Coherence matrix mismatches

In the following analytical insights about the effect of coherence matrix mismatches on the ML and EVD-based PSD estimators are provided. More precisely, it is shown that when coherence matrix mismatches result in an overestimation of the late reverberant PSD, the ML estimator yields a larger estimation error than the EVD-based estimator. Furthermore, when coherence matrix mismatches result in an underestimation of the late reverberant PSD, the ML estimator yields a smaller estimation error than the EVD-based estimator.

Using $\tilde{\Gamma}$ instead of Γ in (12a), the ML late reverberant PSD estimate is given by

$$\hat{\Phi}_r^{\text{ml}} = \frac{1}{M-1} \left(\text{tr} \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\} - \text{tr} \left\{ \frac{\mathbf{d} \mathbf{d}^H \tilde{\Gamma}^{-1}}{\mathbf{d}^H \tilde{\Gamma}^{-1} \mathbf{d}} \Phi_x \tilde{\Gamma}^{-1} \right\} \right). \quad (41)$$

The difference between the ML estimate $\hat{\Phi}_r^{\text{ml}}$ in (41) and the true PSD Φ_r in (17b) is equal to

$$\begin{aligned} \Delta\Phi_r^{\text{ml}} &= \hat{\Phi}_r^{\text{ml}} - \Phi_r \\ &= \frac{1}{M-1} \left(\text{tr} \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\} - \frac{\mathbf{d}^H \tilde{\Gamma}^{-1} \Phi_x \tilde{\Gamma}^{-1} \mathbf{d}}{\mathbf{d}^H \tilde{\Gamma}^{-1} \mathbf{d}} \right. \\ &\quad \left. - \text{tr} \left\{ \Phi_x \Gamma^{-1} \right\} + \lambda_1 \left\{ \Phi_x \Gamma^{-1} \right\} \right), \end{aligned} \quad (42)$$

with the ML estimation error being $|\Delta\Phi_r^{\text{ml}}|$. If $\Delta\Phi_r^{\text{ml}} > 0$ the ML estimator overestimates the true late reverberant PSD whereas if $\Delta\Phi_r^{\text{ml}} < 0$ the ML estimator underestimates the true late reverberant PSD.

Since in the presence of coherence matrix mismatches $\Phi_r \tilde{\Gamma} \tilde{\Gamma}^{-1} \neq \Phi_r \mathbf{I}$, (17) does not hold and different PSD estimates are obtained for the EVD-based estimator depending on the eigenvalue used for the estimation, i.e.,

$$\hat{\Phi}_{r,2}^{\text{evd}} \neq \hat{\Phi}_{r,3}^{\text{evd}} \neq \dots \neq \hat{\Phi}_{r,M}^{\text{evd}} \neq \hat{\Phi}_{r,\text{tr}}^{\text{evd}}, \quad (44)$$

with

$$\hat{\Phi}_{r,i}^{\text{evd}} = \lambda_i \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\}, \quad i = 2, \dots, M, \quad (45)$$

$$\hat{\Phi}_{r,\text{tr}}^{\text{evd}} = \frac{1}{M-1} \left(\text{tr} \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\} - \lambda_1 \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\} \right). \quad (46)$$

The theoretical analysis provided in this section is conducted for the EVD-based PSD estimate $\hat{\Phi}_{r,\text{tr}}^{\text{evd}}$. In the simulation results in Section V, the performance for both $\hat{\Phi}_{r,\text{tr}}^{\text{evd}}$ and $\hat{\Phi}_{r,2}^{\text{evd}}$ is investigated. It should be noted that for a dual-channel scenario, i.e., $M = 2$, $\hat{\Phi}_{r,\text{tr}}^{\text{evd}} = \hat{\Phi}_{r,2}^{\text{evd}}$. The difference between

the EVD-based estimate $\hat{\Phi}_{r,\text{tr}}^{\text{evd}}$ in (46) and the true PSD Φ_r in (17b) is equal to

$$\Delta\Phi_{r,\text{tr}}^{\text{evd}} = \hat{\Phi}_{r,\text{tr}}^{\text{evd}} - \Phi_r \quad (47)$$

$$= \frac{1}{M-1} \left(\text{tr} \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\} - \lambda_1 \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\} \right. \\ \left. - \text{tr} \left\{ \Phi_x \Gamma^{-1} \right\} + \lambda_1 \left\{ \Phi_x \Gamma^{-1} \right\} \right), \quad (48)$$

with the EVD-based estimation error being $|\Delta\Phi_{r,\text{tr}}^{\text{evd}}|$. If $\Delta\Phi_{r,\text{tr}}^{\text{evd}} > 0$ the EVD-based estimator overestimates the true late reverberant PSD whereas if $\Delta\Phi_{r,\text{tr}}^{\text{evd}} < 0$ the EVD-based estimator underestimates the true late reverberant PSD.

In order to compare the estimation errors $|\Delta\Phi_{r,\text{tr}}^{\text{evd}}|$ and $|\Delta\Phi_r^{\text{ml}}|$ in the following it is first shown that $\Delta\Phi_{r,\text{tr}}^{\text{evd}} - \Delta\Phi_r^{\text{ml}} \leq 0$, i.e.,

$$\Delta\Phi_{r,\text{tr}}^{\text{evd}} \leq \Delta\Phi_r^{\text{ml}}. \quad (49)$$

The difference $\Delta\Phi_{r,\text{tr}}^{\text{evd}} - \Delta\Phi_r^{\text{ml}}$ is equal to

$$\Delta\Phi_{r,\text{tr}}^{\text{evd}} - \Delta\Phi_r^{\text{ml}} \quad (50)$$

$$= \frac{1}{M-1} \left(\frac{\mathbf{d}^H \tilde{\Gamma}^{-1} \Phi_x \tilde{\Gamma}^{-1} \mathbf{d}}{\mathbf{d}^H \tilde{\Gamma}^{-1} \mathbf{d}} - \lambda_1 \left\{ \Phi_x \tilde{\Gamma}^{-1} \right\} \right). \quad (51)$$

Let us consider the EVD of $\Phi_x \tilde{\Gamma}^{-1}$, i.e.,

$$\Phi_x \tilde{\Gamma}^{-1} = \mathbf{P} \mathbf{Q} \mathbf{P}^{-1}, \quad (52)$$

with \mathbf{P} an $M \times M$ matrix of eigenvectors and $\mathbf{Q} = \text{diag} \{ [q_1 \ q_2 \ \dots \ q_M]^T \}$ the $M \times M$ diagonal matrix of eigenvalues. Since the matrix Φ_x is a positive definite matrix and the matrix $\tilde{\Gamma}^{-1}$ is a symmetric positive definite matrix, the product $\Phi_x \tilde{\Gamma}^{-1}$ is also positive definite, i.e.,

$$q_1 \geq q_2 \geq \dots \geq q_M > 0. \quad (53)$$

Based on (52) and the fact that the PSD matrix Φ_x is a full rank invertible matrix, the matrix $\tilde{\Gamma}^{-1}$ can be written as

$$\tilde{\Gamma}^{-1} = \Phi_x^{-1} \mathbf{P} \mathbf{Q} \mathbf{P}^{-1}. \quad (54)$$

Combining (52) and (54), (51) can be written as

$$\Delta\Phi_{r,\text{tr}}^{\text{evd}} - \Delta\Phi_r^{\text{ml}} = \frac{1}{M-1} \left(\frac{\mathbf{d}^H \Phi_x^{-1} \mathbf{P} \mathbf{Q}^2 \mathbf{P}^{-1} \mathbf{d}}{\mathbf{d}^H \Phi_x^{-1} \mathbf{P} \mathbf{Q} \mathbf{P}^{-1} \mathbf{d}} - q_1 \right) \quad (55)$$

$$= \frac{1}{M-1} \left(\frac{\mathbf{u}^H \mathbf{Q}^2 \mathbf{v}}{\mathbf{u}^H \mathbf{Q} \mathbf{v}} - q_1 \right), \quad (56)$$

with the vectors $\mathbf{u} = \mathbf{P}^H \Phi_x^{-1} \mathbf{d}$ and $\mathbf{v} = \mathbf{P}^{-1} \mathbf{d}$ introduced in order to simplify the notation. Representing the matrix vector multiplications in (56) as

$$\mathbf{u}^H \mathbf{Q}^2 \mathbf{v} = \sum_{i=1}^M q_i^2 u_i^* v_i, \quad \mathbf{u}^H \mathbf{Q} \mathbf{v} = \sum_{i=1}^M q_i u_i^* v_i, \quad (57)$$

the difference in (56) can be further written as

$$\Delta\Phi_{r,\text{tr}}^{\text{evd}} - \Delta\Phi_r^{\text{ml}} = \frac{1}{M-1} \frac{\sum_{i=1}^M q_i^2 u_i^* v_i - q_1 \sum_{i=1}^M q_i u_i^* v_i}{\sum_{i=1}^M q_i u_i^* v_i} \quad (58)$$

$$= \frac{1}{M-1} \frac{\sum_{i=2}^M (q_i^2 - q_1 q_i) u_i^* v_i}{\sum_{i=1}^M q_i u_i^* v_i}, \quad (59)$$

with $q_i^2 - q_1 q_i \leq 0$ for $i = 1, \dots, M$, cf. (53). In addition, consider that

$$\mathbf{u}^H \mathbf{v} = \mathbf{d}^H \Phi_{\mathbf{x}}^{-1} \mathbf{P} \mathbf{P}^{-1} \mathbf{d} = \mathbf{d}^H \Phi_{\mathbf{x}}^{-1} \mathbf{d} > 0, \quad (60)$$

with the inequality clearly holding since the matrix $\Phi_{\mathbf{x}}^{-1}$ is a symmetric positive definite matrix. Based on (60), it can be said that $u_i^* v_i \geq 0$ for $i = 1, \dots, M$. In summary, since

- $q_i^2 - q_1 q_i \leq 0$,
- $q_i > 0$,
- $u_i^* v_i \geq 0$,

we conclude that

$$\Delta \Phi_{\mathbf{r}, \text{tr}}^{\text{evd}} - \Delta \Phi_{\mathbf{r}}^{\text{ml}} \leq 0 \Rightarrow \Delta \Phi_{\mathbf{r}, \text{tr}}^{\text{evd}} \leq \Delta \Phi_{\mathbf{r}}^{\text{ml}}. \quad (61)$$

Hence, the difference between the EVD-based estimate and the true PSD is always smaller or equal than the difference between the ML estimate and the true PSD. Using (61), the following conclusions can be derived:

- If coherence matrix mismatches result in an overestimation of the true late reverberant PSD for both estimators, i.e., if $\Delta \Phi_{\mathbf{r}, \text{tr}}^{\text{evd}} > 0$ and $\Delta \Phi_{\mathbf{r}}^{\text{ml}} > 0$, it can be said that the EVD-based estimation error is smaller or equal than the ML estimation error, i.e.,

$$|\Delta \Phi_{\mathbf{r}, \text{tr}}^{\text{evd}}| \leq |\Delta \Phi_{\mathbf{r}}^{\text{ml}}|. \quad (62)$$

- If coherence matrix mismatches result in an underestimation of the true late reverberant PSD for both estimators, i.e., $\Delta \Phi_{\mathbf{r}, \text{tr}}^{\text{evd}} < 0$ and $\Delta \Phi_{\mathbf{r}}^{\text{ml}} < 0$, it can be said that the EVD-based estimation error is larger or equal than the ML estimation error, i.e.,

$$|\Delta \Phi_{\mathbf{r}, \text{tr}}^{\text{evd}}| \geq |\Delta \Phi_{\mathbf{r}}^{\text{ml}}|. \quad (63)$$

These theoretical results are experimentally validated in Section V-B.

V. SIMULATION RESULTS

In this section the PSD estimation accuracy of the ML and EVD-based estimators is extensively compared. Furthermore, the performance of a MWF using the considered ML and EVD-based PSD estimators is investigated for different acoustic scenarios. In Section V-A the considered acoustic systems, algorithmic settings, and instrumental measures are introduced. In Section V-B the analytical results of Section IV are experimentally validated. In Sections V-C and V-D the dereverberation performance of a MWF in noiseless scenarios using the considered ML and EVD-based estimators is investigated when the true RETF vector is available and in the presence of RETF vector mismatches. Finally, in Section V-E the dereverberation and noise reduction performance of a MWF in noisy scenarios using the EVD-based estimator is investigated.

As in e.g., [1], [2], the MWF is implemented as an MVDR beamformer \mathbf{w}_{MVDR} followed by a single-channel Wiener post-filter G applied to the MVDR output. For noiseless scenarios as in Sections V-C and V-D, the MWF is implemented as

$$\mathbf{w}_{\text{MWF}} = \underbrace{\tilde{\mathbf{d}}^H \tilde{\mathbf{r}}^{-1} \tilde{\mathbf{d}}}_{\mathbf{w}_{\text{MVDR}}} \underbrace{\frac{\Phi_s}{\Phi_s + \frac{\Phi_r}{\tilde{\mathbf{d}}^H \tilde{\mathbf{r}} \tilde{\mathbf{d}}}}}_G. \quad (64)$$

For noisy scenarios as in Section V-E, the MWF is implemented as

$$\mathbf{w}_{\text{MWF}} = \underbrace{(\Phi_r \tilde{\mathbf{r}} + \Phi_v)^{-1} \tilde{\mathbf{d}}}_{\mathbf{w}_{\text{MVDR}}} \underbrace{\frac{\Phi_s}{\tilde{\mathbf{d}}^H (\Phi_r \tilde{\mathbf{r}} + \Phi_v)^{-1} \tilde{\mathbf{d}} + \tilde{\mathbf{d}}^H (\Phi_r \tilde{\mathbf{r}} + \Phi_v)^{-1} \tilde{\mathbf{d}}}}_G, \quad (65)$$

with the noise component PSD matrix Φ_v estimated from the noise-only signal. It should be noted that independently of the estimator used for the PSDs Φ_s and Φ_r , the MWF implemented according to (64) or (65) is sensitive to mismatches in the RETF vector $\tilde{\mathbf{d}}$ due to the sensitivity of the MVDR beamformer to mismatches in the RETF $\tilde{\mathbf{d}}$. However, as shown in Section V-D, a significantly higher sensitivity of the MWF is observed when the used PSD estimator is also affected by mismatches in $\tilde{\mathbf{d}}$.

A. Acoustic systems, algorithmic settings, and instrumental measures

Acoustic systems. We consider two measured acoustic systems with $M \in \{2, 4, 6, 8\}$ 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 circular microphone array with a radius of 10 cm placed in a room with reverberation time $T_{60} \approx 730$ ms [12]. The second acoustic system consists of a linear microphone array with an inter-sensor distance of 3 cm placed in a room with reverberation time $T_{60} \approx 610$ ms [13]. The sampling frequency is $f_s = 16$ kHz and the reverberant speech components are generated by convolving clean speech signals from the HINT database [14] with the measured RIRs. The noise components consist of a stationary speech shaped directional interference located at an angle $\theta = 345^\circ$ and at a distance of 2 m from the microphone array. The broadband signal to interference ratio is varied between 10 dB and 30 dB. The speech-plus-noise signal is 18 s long and is preceded by a 5 s long noise-only signal such that the noise component PSD matrix can be estimated during speech absence.

Algorithmic settings. The signals are processed using a weighted overlap-add framework with a frame size $N = 1024$ samples and an overlap of 75% between successive frames. The first microphone is arbitrarily selected as the reference microphone. In order to estimate the required PSDs and to implement the MWF, the RETF vector $\tilde{\mathbf{d}}$, the normalized coherence matrix $\tilde{\mathbf{r}}$, the reverberant speech component PSD matrix Φ_s , and the noise component PSD matrix Φ_v are required. The vector $\tilde{\mathbf{d}}$ is computed from the respective RIRs truncated to the part containing only the direct path response [2], [3], [8]. Assuming a spherically isotropic sound field for the reverberation, the normalized spatial coherence between two microphones p and q at a distance d_{pq} is analytically computed as

$$\tilde{r}_{pq}(k) = \text{sinc}\left(\frac{2\pi f_s d_{pq}}{Nc}\right), \quad (66)$$

with c the speed of sound. In none of the considered acoustic scenarios the late reverberation is truly isotropic, resulting

in a mismatch between the available coherence matrix $\tilde{\Gamma}$ and the true coherence matrix Γ . The computation of the reverberant speech component PSD matrix $\Phi_{\mathbf{x}}$ is specified at the beginning of each section. For the noisy scenarios considered in Section V-E, the noise component PSD matrix $\tilde{\Phi}_{\mathbf{v}}$ is computed as

$$\Phi_{\mathbf{v}}(k) = \frac{1}{L_v} \sum_{l=0}^{L_v-1} \mathbf{v}(k, l) \mathbf{v}^H(k, l), \quad (67)$$

with L_v the total number of noise-only frames. The minimum gain of the single-channel Wiener postfilter is set to -20 dB.

Instrumental measures. In order to validate the theoretical results in Section IV, the estimated late reverberant PSD $\hat{\Phi}_{\mathbf{r}}$ is compared to the true PSD $\Phi_{\mathbf{r}}$ in terms of the overestimation and underestimation error measures. The overestimation error OvErr is computed as

$$\text{OvErr} = \frac{1}{LN} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} \left| \min \left(0, 10 \log_{10} \frac{\Phi_{\mathbf{r}}(k, l)}{\hat{\Phi}_{\mathbf{r}}(k, l)} \right) \right|, \quad (68)$$

with K the frame size. The underestimation error UnErr is computed as

$$\text{UnErr} = \frac{1}{LN} \sum_{l=0}^{L-1} \sum_{k=0}^{N-1} \left| \max \left(0, 10 \log_{10} \frac{\Phi_{\mathbf{r}}(k, l)}{\hat{\Phi}_{\mathbf{r}}(k, l)} \right) \right|. \quad (69)$$

The total estimation error ToErr is then given by

$$\text{ToErr} = \text{OvErr} + \text{UnErr}. \quad (70)$$

Furthermore the instrumental measures used in Sections V-C-V-E to evaluate the dereverberation and noise reduction performance are the frequency weighted segmental signal-to-noise ratio (fwSSNR) [15], the cepstral distance (CD) measure [16], and the perceptual evaluation of speech quality (PESQ) measure [17]. In [12] it has been shown that measures such as fwSSNR, CD, and PESQ can correlate well with the perceived overall quality of signals processed by a wide range of dereverberation and noise reduction techniques. These measures are intrusive measures, generating a similarity score between a test signal and a reference signal. The reference signal employed here is the clean speech signal. The improvement in fwSSNR, CD, and PESQ, i.e., ΔfwSSNR , ΔCD , and ΔPESQ , is computed as the difference between the fwSSNR, CD, and PESQ scores of the output signal and the first microphone signal. Note that a positive ΔfwSSNR , a negative ΔCD , and a positive ΔPESQ indicate a performance improvement.

B. Theoretical results validation

In this section the theoretical results derived in Section IV are experimentally validated using acoustic system S_1 and $M = 4$ microphones. In order to avoid any other sources of error, the reverberant speech component PSD matrix $\Phi_{\mathbf{x}}(k, l)$ is set to

$$\Phi_{\mathbf{x}}(k, l) = \Phi_{\mathbf{s}}(k, l) \mathbf{d}(k) \mathbf{d}^H(k) + \Phi_{\mathbf{r}}(k, l) \Gamma(k), \quad (71)$$

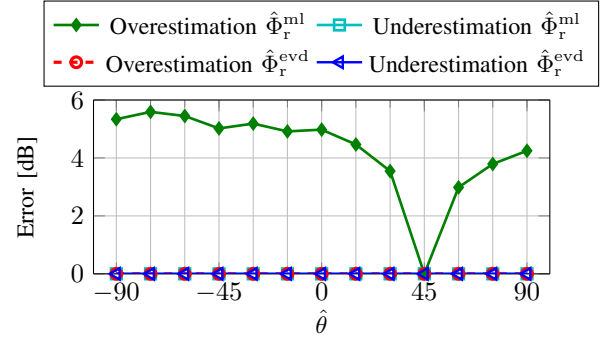


Fig. 2: Overestimation and underestimation error using the ML and EVD-based estimators in the presence of RETF vector mismatches (acoustic system S_1 , $M = 4$, $\theta = 45^\circ$).

with \mathbf{d} computed from the truncated RIRs corresponding to the true DOA $\theta = 45^\circ$ and $\Gamma(k)$ the spherically isotropic coherence matrix. The PSDs $\Phi_{\mathbf{s}}(k, l)$ and $\Phi_{\mathbf{r}}(k, l)$ are computed using recursive averaging as

$$\Phi_{\mathbf{s}}(k, l) = \alpha \Phi_{\mathbf{s}}(k, l-1) + (1-\alpha) |S(k, l)|^2, \quad (72)$$

$$\Phi_{\mathbf{r}}(k, l) = \alpha \Phi_{\mathbf{r}}(k, l-1) + (1-\alpha) |X_{\mathbf{r},1}(k, l)|^2, \quad (73)$$

with $\alpha = 0.67$.

PSD estimation accuracy in the presence of RETF vector mismatches. In order to evaluate the PSD estimation accuracy of the ML and EVD-based estimators in the presence of RETF vector mismatches, in this section the available RETF vector $\tilde{\mathbf{d}}$ is computed from the truncated RIRs corresponding to several erroneous DOAs $\hat{\theta} \in \{-90^\circ, -75^\circ, \dots, 90^\circ\}$. Using $\Phi_{\mathbf{x}}(k, l)$, $\Gamma(k)$, and $\tilde{\mathbf{d}}(k)$ the late reverberant PSD $\Phi_{\mathbf{r}}(k, l)$ is estimated using the ML and EVD-based estimators and the PSD estimation accuracy is evaluated using the OvErr and the UnErr measures in (68) and (69). Fig. 2 illustrates the overestimation and underestimation error obtained when using the ML and EVD-based estimators to estimate the late reverberant PSD in the presence of RETF vector mismatches. As analytically shown in Section IV, it can be confirmed that in the presence of RETF vector mismatches, the ML estimator always overestimates the late reverberant PSD. Furthermore, it can be observed that for perfect knowledge of the RETF vector (i.e., $\hat{\theta} = \theta = 45^\circ$) the ML estimator yields the true late reverberant PSD, as indicated by the fact that the overestimation and underestimation errors are both 0. As expected, since the EVD-based estimator is independent of the RETF vector, the overestimation and underestimation error are both 0, i.e., the proposed EVD-based estimator always yields the true late reverberant PSD.

PSD estimation accuracy in the presence of coherence matrix mismatches.

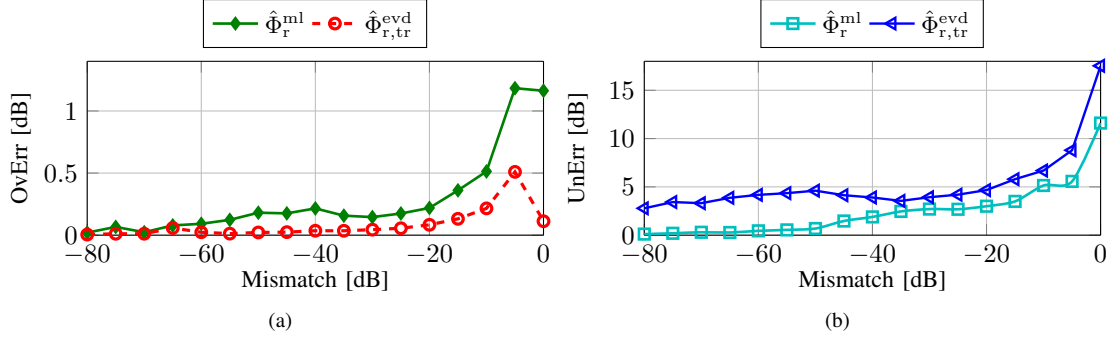


Fig. 3: Performance of the MVDR beamformer and the MWF using $\hat{\Phi}_r^{ml}$, $\hat{\Phi}_{r,tr}^{evd}$, and $\hat{\Phi}_{r,2}^{evd}$ for acoustic system 1 and perfectly estimated DOA in terms of (a) $\Delta fwSSNR$, (b) ΔCD , and (c) $\Delta PESQ$.

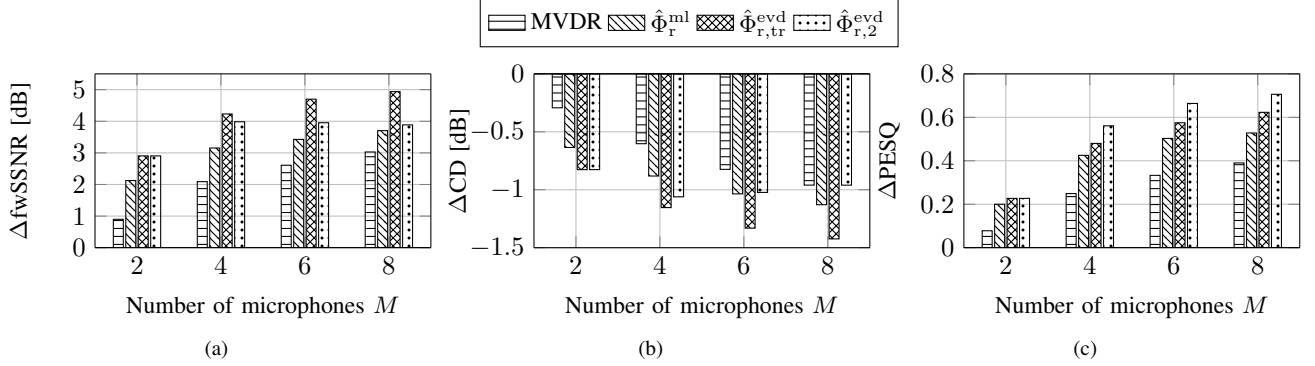


Fig. 4: Performance of the MVDR beamformer and the MWF using $\hat{\Phi}_r^{ml}$, $\hat{\Phi}_{r,tr}^{evd}$, and $\hat{\Phi}_{r,2}^{evd}$ for acoustic system 1 and perfectly estimated DOA in terms of (a) $\Delta fwSSNR$, (b) ΔCD , and (c) $\Delta PESQ$.

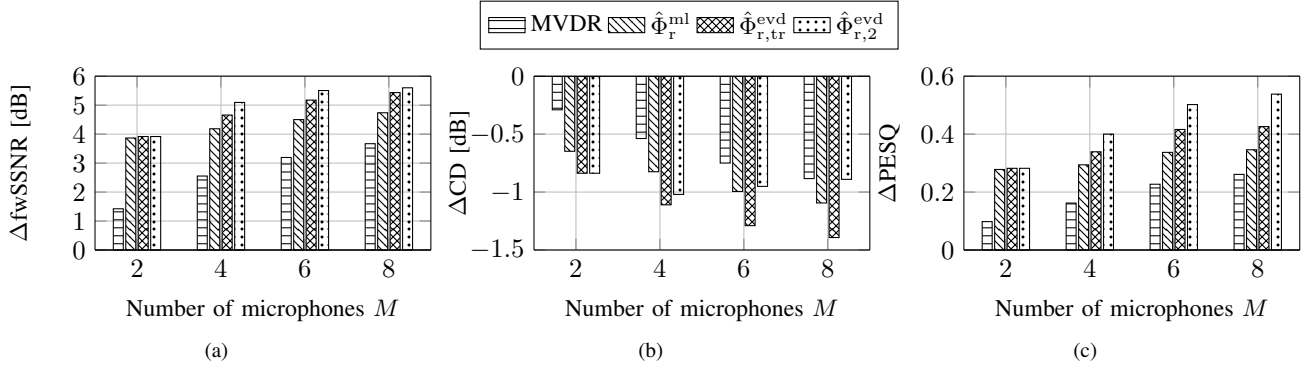


Fig. 5: Performance of the MVDR beamformer and the MWF using $\hat{\Phi}_r^{ml}$, $\hat{\Phi}_{r,tr}^{evd}$, and $\hat{\Phi}_{r,2}^{evd}$ for acoustic system 2 and perfectly estimated DOA in terms of (a) $\Delta fwSSNR$, (b) ΔCD , and (c) $\Delta PESQ$.

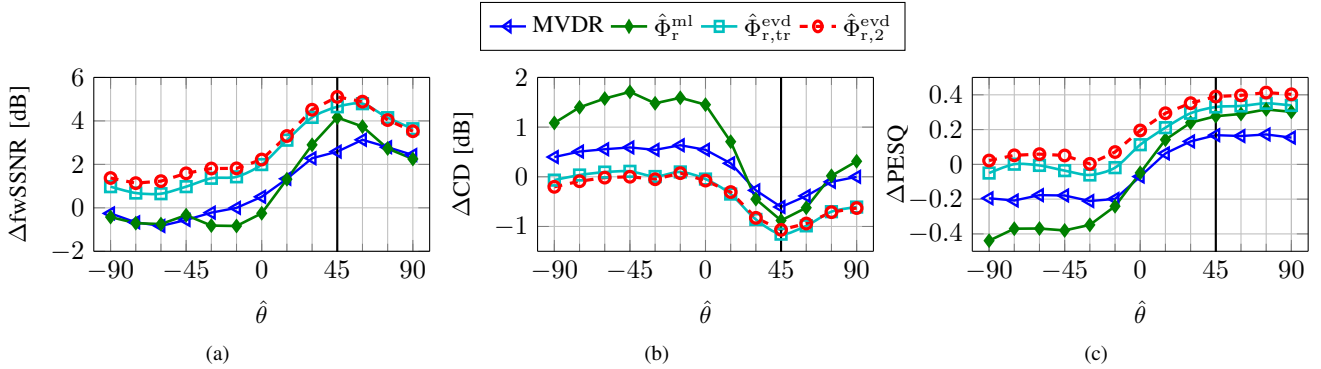


Fig. 6: Performance of the MVDR beamformer and the MWF using $\hat{\Phi}_r^{ml}$, $\hat{\Phi}_{r,tr}^{evd}$, and $\hat{\Phi}_{r,2}^{evd}$ for acoustic system 2 in the presence of DOA estimation errors in terms of (a) $\Delta fwSSNR$, (b) ΔCD , and (c) $\Delta PESQ$ ($M = 4$, $\theta = 45^\circ$).

TABLE I: Dereverberation performance of the MWF using $\hat{\Phi}_{r,tr}^{evd}$ and $\hat{\Phi}_{r,2}^{evd}$ for acoustic system 2 and perfectly estimated DOA in the presence of additive noise ($M = 4$).

Input SNR [dB]	$\Delta fwSSNR$ [dB]		ΔCD [dB]	
	$\hat{\Phi}_{r,tr}^{evd}$	$\hat{\Phi}_{r,2}^{evd}$	$\hat{\Phi}_{r,tr}^{evd}$	$\hat{\Phi}_{r,2}^{evd}$
10	3.95	4.77	-0.80	-0.83
20	4.32	5.00	-1.00	-0.93
30	4.50	4.93	-1.04	-0.92

TABLE II: Dereverberation and noise reduction performance of the MWF using $\hat{\Phi}_{r,tr}^{evd}$ and $\hat{\Phi}_{r,2}^{evd}$ for acoustic system 2 and perfectly estimated DOA in the presence of additive noise ($M = 4$).

Input SNR [dB]	$\Delta fwSSNR$ [dB]		ΔCD [dB]	
	$\hat{\Phi}_{r,tr}^{evd}$	$\hat{\Phi}_{r,2}^{evd}$	$\hat{\Phi}_{r,tr}^{evd}$	$\hat{\Phi}_{r,2}^{evd}$
10	8.08	9.35	-1.65	-1.63
20	5.97	6.70	-1.23	-1.14
30	4.92	5.36	-1.05	-0.93

C. Dereverberation performance when the true RETF vector is available

D. Dereverberation performance in the presence of RETF vector mismatches

E. Dereverberation and noise reduction performance

REFERENCES

- [1] 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, Sep. 2013.
- [2] A. Kuklasinski, 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, Sep. 2014.
- [3] A. Kuklasinski, 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] A. Kuklasinski, 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, Sep. 2016.
- [8] —, "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.
- [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] 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, Jun. 2015.
- [11] 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.
- [12] 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.
- [13] 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, Sep. 2014, pp. 313–317.
- [14] 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.
- [15] 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.
- [16] S. Quackenbush, T. Barnwell, and M. Clements, *Objective measures of speech quality*. New Jersey, USA: Prentice-Hall, 1988.
- [17] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.