MULTI-CHANNEL LATE REVERBERATION POWER SPECTRAL DENSITY ESTIMATION BASED ON NUCLEAR NORM MINIMIZATION

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 reverberation power spectral density (PSD) generally assume that the reverberant PSD matrix can be decomposed as the sum of a rank-1 matrix and a scaled diffuse coherence matrix. To account for modeling or estimation errors in the estimated reverberant PSD matrix, in this paper we propose to decompose this matrix as the sum of a low rank (not necessarily rank-1) matrix and a scaled diffuse coherence matrix. Among all pairs of scalars and matrices that yield feasible decompositions, the late reverberation PSD can then be estimated as the scalar associated with the matrix of minimum rank. Since rank minimization is an intractable non-convex optimization problem, we propose to use a convex relaxation approach and estimate the late reverberation PSD based on nuclear norm minimization (NNM). Experimental results show the advantages of using the proposed NNM-based late reverberation PSD estimator in a multi-channel Wiener filter for speech dereverberation, significantly outperforming a state-of-the-art maximum likelihood-based PSD estimator and yielding a similar or better performance than a recently proposed eigenvalue decomposition-based PSD estimator.

Index Terms— dereverberation, nuclear norm, convex optimization, MWF, PSD estimation

1. INTRODUCTION

In hands-free communication applications the recorded microphone signals are often corrupted by early and late reverberation, which arises from the superposition of delayed and attenuated copies of the anechoic speech signal. While early reverberation may be desirable [1], late reverberation may degrade the perceived quality and hinder the intelligibility of speech [2]. Hence, speech enhancement techniques which effectively suppress the late reverberation are required. In the last decades many single-channel and multi-channel dereverberation techniques have been proposed [3], with multi-channel techniques being generally preferred since they are able to exploit both the spectro-temporal and the spatial characteristics of the received microphone signals. Many such techniques require an estimate of the late reverberation power spectral density (PSD), e.g. [4–6].

The late reverberation PSD can be estimated either using single-channel estimators based on a temporal model of reverberation [7,8] or multi-channel estimators based on a (spatial) diffuse sound field model of reverberation [9–17]. Most multi-channel PSD estimators [9–15] require an estimate of the relative early transfer functions (RETFs) of the target signal from the reference microphone to all microphones, which may be difficult to accurately estimate, particularly in highly reverberant and noisy scenarios. Recently, we proposed a multi-channel late reverberation PSD estimator based on an eigenvalue decomposition (EVD), which does not require such RETF estimates [16, 17]. Experimental results in [17] show the advantages of using this EVD-based estimator in a multi-channel Wiener filter (MWF) for speech dereverberation, outperforming the maximum likelihood (ML)-based estimator in [10] both when the RETFs are perfectly estimated as well as in the presence of RETF estimation errors.

The EVD-based estimator in [17] relies on the assumption that the reverberant PSD matrix is equal to the sum of a rank-1 matrix (corresponding to the direct and early reverberation speech component) and a diffuse coherence matrix scaled with the late reverberation PSD. However, since the late reverberation is not truly diffuse and since the reverberant PSD matrix in practice is estimated using a signal realization, the estimated reverberant PSD matrix can deviate from this assumption. In order to account for this deviation, in this paper we propose to model the reverberant PSD matrix as the sum of a low rank (not necessarily rank-1) matrix and a scaled diffuse coherence matrix. Among all pairs of scalars and matrices that yield feasible decompositions, the late reverberation PSD can be estimated as the scalar associated with the matrix of minimum rank. However, since the rank of a matrix is non-convex and non-convex optimization problems are typically hard (if not impossible) to solve, we propose to estimate the late reverberation PSD based on nuclear norm minimization (NNM) [18–20] instead. The nuclear norm is a convex relaxation of the rank, and hence, NNM-based optimization problems can be efficiently solved [18]. Experimental results for several acoustic systems and configurations illustrate the advantages of using the NNM-based PSD estimator in an MWF for speech dereverberation, yielding a similar or better performance than the ML-based and EVD-based PSD estimators.

2. 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. In the short-time Fourier transform (STFT) domain, the M-dimensional vector of the microphone signals $\mathbf{y}(k,l) = [Y_1(k,l) \ \dots \ Y_M(k,l)]^T$ at frequency index k and frame index l is given by

$$\mathbf{y}(k,l) = \underbrace{\mathbf{x}_{\mathrm{e}}(k,l) + \mathbf{x}_{\mathrm{r}}(k,l)}_{\mathbf{x}(k,l)} + \mathbf{v}(k,l), \tag{1}$$

with $\mathbf{x}(k,l)$ the speech component, $\mathbf{v}(k,l)$ the noise component, $\mathbf{x}_{\mathrm{e}}(k,l)$ the direct and early reverberation speech component, and $\mathbf{x}_{\mathrm{r}}(k,l)$ the late reverberation speech component. For simplicity, in the following we assume that the noise component is equal to zero, i.e., $\mathbf{y}(k,l) = \mathbf{x}(k,l)$. However, the late reverberation PSD

This work was supported in part by the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG), and the joint Lower Saxony-Israeli Project ATHENA, funded by the State of Lower Saxony.

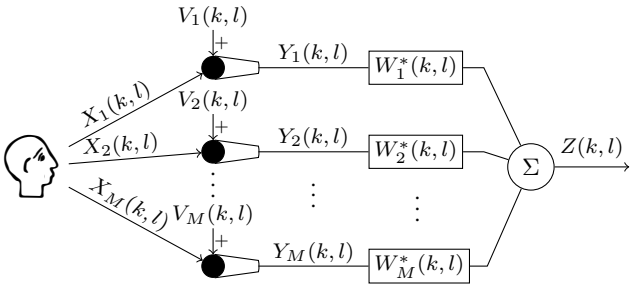

Figure 1: Acoustic system configuration.

estimator proposed in this paper can also be used in noisy scenarios, cf. Section 3.3.

The direct and early reverberation speech component $\mathbf{x}_{\mathrm{e}}(k,l)$ can be expressed as

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

with $S(k,l)$ the target signal (i.e., direct and early reverberation speech component) received by the reference microphone and $\mathbf{d}(k) = [D_1(k) \ \dots \ D_M(k)]^T$ the vector of RETFs of the target signal from the reference microphone to all microphones. The late reverberation speech component $\mathbf{x}_{\mathrm{r}}(k,l)$ is commonly modeled as a diffuse sound component and is assumed to be uncorrelated with the direct and early reverberation speech component $\mathbf{x}_{\mathrm{e}}(k,l)$ [9–17]. Hence, the reverberant PSD matrix can be written as

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{x}}(k,l) &= \mathcal{E}\{\mathbf{x}(k,l)\mathbf{x}^H(k,l)\} &(3)\\ &= \mathcal{E}\{\mathbf{x}_{\mathrm{e}}(k,l)\mathbf{x}_{\mathrm{e}}^H(k,l)\} + \mathcal{E}\{\mathbf{x}_{\mathrm{r}}(k,l)\mathbf{x}_{\mathrm{r}}^H(k,l)\}, &(4)\end{aligned}$$

with $\mathcal{E}$ the expectation operator. Based on (2) and on a diffuse sound field model for the late reverberation, the PSD matrix $\mathbf{\Phi}_{\mathbf{x}}(k,l)$ can be expressed as the sum of a rank-1 matrix and a scaled diffuse coherence matrix, i.e.,

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

with $\Phi_{\mathrm{s}}(k,l) = \mathcal{E}\{|S(k,l)|^2\}$ the (time-varying) PSD of the target signal, $\Phi_{\mathrm{r}}(k,l)$ the (time-varying) PSD of the late reverberation, and $\mathbf{\Gamma}(k)$ the (time-invariant) coherence matrix of a diffuse sound field, which can be analytically computed based on the microphone array geometry [21]. In practice, an estimate of the PSD matrix $\mathbf{\Phi}_{\mathbf{x}}(k,l)$ is obtained using recursive averaging with a smoothing factor α, i.e.,

$$\hat{\mathbf{\Phi}}_{\mathbf{x}}(k,l) = \alpha\mathbf{x}(k,l)\mathbf{x}^H(k,l) + (1-\alpha)\hat{\mathbf{\Phi}}_{\mathbf{x}}(k,l-1). \tag{6}$$

Given the filter vector $\mathbf{w}(k,l) = [W_1(k,l) \ \dots \ W_M(k,l)]^T$, the output signal $Z(k,l)$ of the speech enhancement system in Fig. 1 can be computed as

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

Speech dereverberation techniques aim at designing the filter $\mathbf{w}(k,l)$ such that the output signal $Z(k,l)$ is as close as possible to the target signal $S(k,l)$. Many such techniques require an estimate of the late reverberation PSD $\Phi_{\mathrm{r}}(k,l)$, e.g. [4–6].

3. LATE REVERBERATION PSD ESTIMATOR

In this section, the ML-based estimator [10] and the EVD-based estimator [17] are briefly reviewed and a novel nuclear norm minimization-based estimator is proposed. Since the estimation is performed independently in each frequency bin, the frequency index k will be omitted in the remainder of this paper.

3.1. Maximum likelihood-based estimator

In order to derive the ML-based estimator in [10], the early and late reverberation speech components are assumed to be circularly-symmetric complex Gaussian distributed. These distributions are then used to construct and maximize a likelihood function, yielding the ML-based late reverberation PSD estimate

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

where $\mathbf{I}$ denotes the $M \times M$-dimensional identity matrix and $\mathrm{tr}\{\cdot\}$ denotes the matrix trace operator. Note that the PSD estimate in (8) requires knowledge of the RETF vector $\mathbf{d}$, which may be difficult to estimate accurately.

3.2. Eigenvalue decomposition-based estimator

To remove the dependency of the PSD estimate on the RETF vector $\mathbf{d}$, we recently proposed to estimate the late reverberation PSD using the EVD of the prewhitened reverberant PSD matrix $\mathbf{\Gamma}^{-1}\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)$ [17]. Based on the model in (5), the EVD-based late reverberation PSD estimate is computed as

$$\hat{\Phi}_{\mathrm{r}}^{\mathrm{evd}}(l) = \frac{1}{M-1}\left(\mathrm{tr}\{\mathbf{\Gamma}^{-1}\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)\} - \lambda_{\max}\{\mathbf{\Gamma}^{-1}\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)\}\right), \tag{9}$$

where $\lambda_{\max}\{\mathbf{\Gamma}^{-1}\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)\}$ denotes the maximum eigenvalue of the prewhitened reverberant PSD matrix. Unlike the ML-based estimate in (8), the EVD-based estimate in (9) does not require knowledge of the RETF vector $\mathbf{d}$, which is advantageous in order to avoid propagation of RETF estimation errors into the PSD estimate. As has been experimentally validated in [17], using the EVD-based PSD estimate in an MWF yields a better dereverberation performance than the ML-based estimate, both for perfectly estimated RETFs as well as in the presence of RETF estimation errors.

3.3. Nuclear norm minimization-based estimator

The EVD-based PSD estimator in (9) relies on the assumptions that 1) the late reverberation can be modeled as a diffuse sound field, 2) the components $\mathbf{x}_{\mathrm{e}}(l)$ and $\mathbf{x}_{\mathrm{r}}(l)$ are uncorrelated, and 3) the estimated PSD matrix $\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)$ in (6) is equal to the PSD matrix $\mathbf{\Phi}_{\mathbf{x}}(l)$ in (5). However, the late reverberation is not perfectly diffuse. Furthermore, even if the components $\mathbf{x}_{\mathrm{e}}(l)$ and $\mathbf{x}_{\mathrm{r}}(l)$ were truly uncorrelated, the PSD matrix $\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)$ in (6) estimated using a realization of $\mathbf{x}(l)$ will likely contain non-zero contributions of the cross-terms $\mathbf{x}_{\mathrm{e}}(l)\mathbf{x}_{\mathrm{r}}^H(l)$ and $\mathbf{x}_{\mathrm{r}}(l)\mathbf{x}_{\mathrm{e}}^H(l)$. As a result, in practice $\hat{\mathbf{\Phi}}_{\mathrm{x}}(l)$ differs from $\mathbf{\Phi}_{\mathrm{x}}(l)$, i.e.,

$$\hat{\mathbf{\Phi}}_{\mathrm{x}}(l) = \mathbf{E}(l) + \mathbf{\Phi}_{\mathrm{x}}(l) = \underbrace{\mathbf{E}(l) + \Phi_{\mathrm{s}}(l)\mathbf{d}\mathbf{d}^H}_{\mathbf{\Delta}(l)} + \Phi_{\mathrm{r}}(l)\mathbf{\Gamma}, \tag{10}$$

with $\mathbf{E}(l)$ an $M \times M$-dimensional Hermitian error matrix and the matrix $\mathbf{\Delta}(l) = \mathbf{E}(l) + \Phi_{\mathrm{s}}(l)\mathbf{d}\mathbf{d}^H$ defined to simplify the notation. Assuming that the matrix $\mathbf{\Delta}(l)$ can be modeled as a low rank (however, not necessarily rank-1) matrix, we propose to estimate the late reverberation PSD $\Phi_{\mathrm{r}}(l)$ by decomposing the estimated PSD matrix $\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)$ into the sum of an unknown low rank Hermitian matrix and

a scaled diffuse coherence matrix. This corresponds to solving the constrained minimization problem

$$\min_{\Phi_{\mathrm{r}}(l),\mathbf{\Delta}(l)} \mathcal{R}\{\mathbf{\Delta}(l)\} \text{ subject to } \begin{cases} \hat{\mathbf{\Phi}}_{\mathbf{x}}(l) = \mathbf{\Delta}(l) + \Phi_{\mathrm{r}}(l)\mathbf{\Gamma}, \\ \Phi_{\mathrm{r}}(l) \geq 0, \\ \mathbf{\Delta}(l) = \mathbf{\Delta}^H(l), \end{cases} \tag{11}$$

where $\mathcal{R}\{\mathbf{\Delta}(l)\}$ denotes the rank of $\mathbf{\Delta}(l)$, defined as the number of nonzero singular values $\sigma_p\{\mathbf{\Delta}(l)\}$, $p = 1, \ldots, M$. Rank minimization problems arise in many statistical modeling and signal processing applications such as in robust principal component analysis [19] and subspace segmentation [20]. However, the matrix rank is non-convex and it is well known that non-convex optimization problems are typically hard (if not impossible) to solve. A common alternative to rank minimization problems is to use a convex relaxation approach and replace the non-convex rank $\mathcal{R}\{\mathbf{\Delta}(l)\}$ with the convex nuclear norm $\|\mathbf{\Delta}(l)\|_*$ [18–20], defined as

$$\|\mathbf{\Delta}(l)\|_* = \sum_{p=1}^{M} \sigma_p\{\mathbf{\Delta}(l)\}. \tag{12}$$

Whereas the rank counts the number of nonzero singular values, the nuclear norm sums the amplitude of the singular values, and it can be shown that under certain conditions low rank solutions can be perfectly recovered via nuclear norm minimization [22]. Hence, we propose to estimate the late reverberation PSD by solving the nuclear norm minimization problem

$$\hat{\Phi}_{\mathrm{r}}^{\mathrm{nnm}}(l) = \underset{\Phi_{\mathrm{r}}(l),\mathbf{\Delta}(l)}{\arg\min} \|\mathbf{\Delta}(l)\|_* \text{ subject to } \begin{cases} \hat{\mathbf{\Phi}}_{\mathbf{x}}(l) = \mathbf{\Delta}(l) + \Phi_{\mathrm{r}}(l)\mathbf{\Gamma}, \\ \Phi_{\mathrm{r}}(l) \geq 0, \\ \mathbf{\Delta}(l) = \mathbf{\Delta}^H(l). \end{cases} \tag{13}$$

Since the optimization problem in (13) is convex, it can be efficiently solved using existing optimization tools, e.g. the Matlab software CVX [23].

It should be noted that although a noise-free scenario is assumed in this paper, the proposed NNM-based estimator can also be used in a noisy scenario, as long as an estimate of the PSD matrix $\mathbf{\Phi}_{\mathbf{x}}(l)$ can be obtained. An estimate of $\mathbf{\Phi}_{\mathbf{x}}(l)$ can in practice be computed by, e.g., subtracting an estimate of the noise PSD matrix from the noisy signal PSD matrix. However, if the noise can also be modeled as a diffuse sound field, the NNM-based estimator can be readily used to estimate the joint late reverberation and noise PSD.

4. EXPERIMENTAL RESULTS

In this section, the dereverberation performance of an MWF using the proposed NNM-based PSD estimator is investigated and compared to the ML-based [10] and EVD-based [17] PSD estimators, both for perfectly as well as for erroneously estimated RETFs. The MWF is implemented as an MVDR beamformer $\mathbf{w}_{\text{MVDR}}$ followed by a single-channel Wiener postfilter $G(l)$, i.e.,

$$\mathbf{w}_{\text{MWF}}(l) = \underbrace{\frac{\mathbf{\Gamma}^{-1}\mathbf{d}}{\mathbf{d}^H\mathbf{\Gamma}^{-1}\mathbf{d}}}_{\mathbf{w}_{\text{MVDR}}} \underbrace{\frac{\hat{\Phi}_{\mathrm{s}}(l)}{\hat{\Phi}_{\mathrm{s}}(l) + \frac{\hat{\Phi}_{\mathrm{r}}(l)}{\mathbf{d}^H\mathbf{\Gamma}^{-1}\mathbf{d}}}}_{G(l)}, \tag{14}$$

with $\hat{\Phi}_{\mathrm{s}}(l)$ and $\hat{\Phi}_{\mathrm{r}}(l)$ the estimated target signal and late reverberation PSDs. When using the ML-based late reverberation PSD estimate $\hat{\Phi}_{\mathrm{r}}^{\mathrm{ml}}(l)$, the target signal PSD $\hat{\Phi}_{\mathrm{s}}(l)$ is estimated within the ML framework as proposed in [10], whereas when using the EVD-based and NNM-based late reverberation PSD estimates $\hat{\Phi}_{\mathrm{r}}^{\mathrm{evd}}(l)$ and $\hat{\Phi}_{\mathrm{r}}^{\mathrm{nnm}}(l)$ respectively, the target signal $\hat{\Phi}_{\mathrm{s}}(l)$ is estimated using the decision directed approach [24]. It should be noted that independently of the late reverberation PSD estimator used, the MWF implemented according to (14) is sensitive to estimation errors in the RETF vector $\mathbf{d}$ due to the sensitivity of the MVDR beamformer to RETF errors. However, as will be illustrated in Section 4.3, a significantly higher sensitivity of the MWF is observed when the late reverberation PSD estimator is also affected by RETF errors.

4.1. Setup

We consider two multi-channel acoustic systems with a single speech source and $M \in \{2, 4, 6\}$ microphones. The first acoustic system consists of a uniform linear microphone array with an inter-microphone distance of 8 cm, placed in a room with reverberation time $T_{60} \approx 0.61$ s [25]. The speech source is located at an angle $\theta = 45^\circ$ and distance $d_{\mathrm{sm}} = 2$ m from the microphone array. The second acoustic system consists of a uniform linear microphone array with an inter-microphone distance of 6 cm, placed in a room with reverberation time $T_{60} \approx 1.25$ ms [26]. The speech source is located at an angle $\theta = -65^\circ$ and distance $d_{\mathrm{sm}} = 2.1$ m from the microphone array. The sampling frequency is $f_s = 16$ kHz and the received reverberant signals are generated by convolving clean speech signals from the HINT database [27] with measured RIRs.

The signals are processed using a weighted overlap-add STFT 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. The RETF vector $\mathbf{d}$ is computed from the truncated RIRs containing only the direct path and early reflections (up to 10 ms). The diffuse coherence matrix $\mathbf{\Gamma}$ is computed based on the microphone array geometry, assuming a spherically diffuse sound field. To estimate the reverberant PSD matrix $\hat{\mathbf{\Phi}}_{\mathbf{x}}(l)$, recursive averaging with a smoothing factor α corresponding to a time constant of 40 ms is used, cf. (6). The minimum gain of the single-channel Wiener postfilter $G(l)$ in (14) is -20 dB.

The performance is evaluated in terms of the improvement in PESQ (ΔPESQ) [28] and cepstral distance (ΔCD) [29] between the output signal and the reference microphone signal. The PESQ and CD measures are intrusive measures comparing the signal being evaluated to a reference signal. The reference signal used in this paper is the anechoic speech signal. It should be noted that a positive ΔPESQ and a negative ΔCD indicate a performance improvement.

The performance of the MVDR beamformer and the MWF implemented according to (14) using the ML-, EVD-, and proposed NNM-based PSD estimators is investigated for

i) both acoustic systems with different number of microphones $M \in \{2, 4, 6\}$ assuming *perfectly estimated RETFs*, i.e., $\mathbf{d}$ is computed from the truncated RIRs measured for the true direction of arrival (DOA) θ of the speech source (Section 4.2),

ii) the first acoustic system with $M = 4$ microphones assuming *erroneously estimated RETFs*, i.e., $\mathbf{d}$ is computed from the truncated RIRs measured for DOAs $\hat{\theta}$ which differ from the true DOA θ (Section 4.3).

4.2. Perfectly estimated RETFs

In this section, the performance of the MVDR beamformer and the MWF using the considered late reverberation PSD estimators is investigated for perfectly estimated RETFs. Fig. 2 depicts the ΔPESQ and ΔCD obtained for all considered acoustic systems and configurations. As expected, it can be observed that for all acoustic systems and configurations, the MWF using any of the considered PSD estimators improves the performance in comparison to the MVDR beamformer in terms of both instrumental measures. In

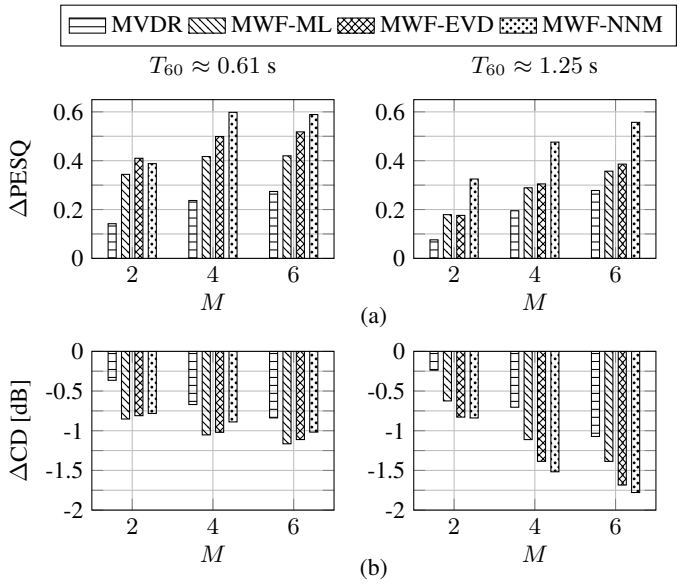

Figure 2: Performance of the MVDR beamformer and the MWF using different late reverberation PSD estimators with perfectly estimated RETFs: (a) ΔPESQ and (b) ΔCD.

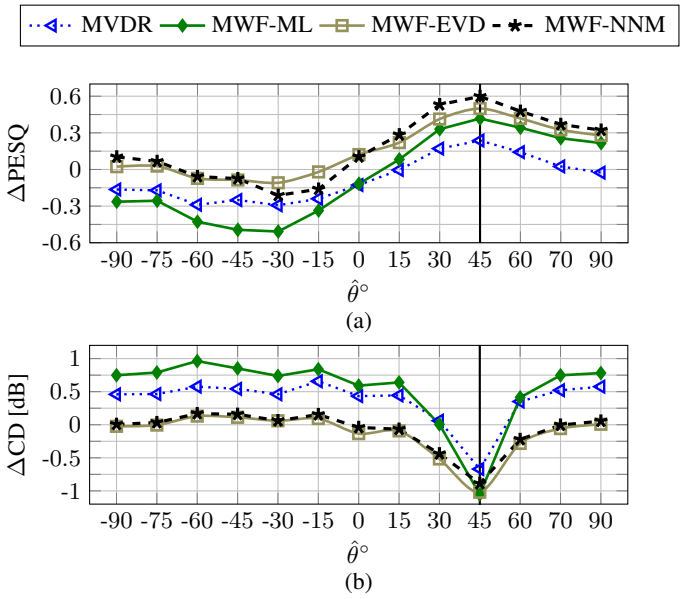

Figure 3: Performance of the MVDR beamformer and the MWF using different late reverberation PSD estimators with erroneously estimated RETFs: (a) ΔPESQ and (b) ΔCD ($T_{60} \approx 0.61$ s, $M = 4$).

terms of ΔPESQ, Fig. 2(a) shows that the proposed NNM-based PSD estimator typically results in the best performance (except for $T_{60} \approx 0.61$ s and $M = 2$ microphones), yielding a ΔPESQ increase of up to 0.2 in comparison to the ML-based and EVD-based PSD estimators. In terms of ΔCD, Fig. 2(b) shows that for the first acoustic system all considered PSD estimators yield a similar performance. For the second acoustic system, Fig. 2(b) shows that the NNM-based PSD estimator outperforms the ML-based PSD estimator and results in a similar or slightly better performance than the EVD-based PSD estimator. Informal listening tests suggest that the NNM-based PSD estimator typically yields a larger suppression of the late reverberation, introducing as a result slightly more signal distortions than the ML-based and EVD-based PSD estimators.

In summary, instrumental measures show that the proposed NNM-based PSD estimator generally yields a better performance than the ML-based and EVD-based PSD estimators when used in an MWF with perfectly estimated RETFs.

4.3. Erroneously estimated RETFs

In this section, the performance of the MVDR beamformer and the MWF using the considered late reverberation PSD estimators is investigated for erroneously estimated RETFs. Fig. 3 depicts the ΔPESQ and ΔCD obtained for the first acoustic system and $M = 4$ microphones when the RETF vector $\mathbf{d}$ is computed from the truncated RIRs measured for several erroneous DOAs. For completeness, the performance obtained for the perfectly estimated RETF vector $\mathbf{d}$ (i.e., $\hat{\theta} = 45°$) is also depicted. As expected, Fig. 3 shows that the performance of the MVDR beamformer deteriorates in the presence of RETF estimation errors in terms of both instrumental measures. Since the MWF is equivalent to an MVDR beamformer followed by a single-channel Wiener postfilter, cf. (14), it can be observed that RETF estimation errors yield a performance deterioration also for the MWF using any of the considered PSD estimators. However, since the ML-based PSD estimator additionally relies on the RETF vector, Fig. 3 shows that the ML-based PSD estimator even worsens the performance in comparison to the MVDR beamformer, resulting in a significantly faster and larger performance deterioration than the EVD-based or the proposed NNM-based PSD estimators. When comparing the EVD-based and NNM-based PSD estimators, Fig. 3(a) shows that in terms of ΔPESQ, the proposed NNM-based PSD estimator typically results in a similar or slightly better performance than the EVD-based estimator (except for $\hat{\theta} = -30°$ and $\hat{\theta} = -15°$). Fig. 3(b) shows that in terms of ΔCD, the proposed NNM-based PSD estimator yields a very similar performance as the EVD-based PSD estimator. Informal listening test suggest that in the presence of RETF estimation errors, using the NNM-based PSD estimator in an MWF yields a larger suppression of the late reverberation than using the EVD-based PSD estimator.

In summary, instrumental measures show that the proposed NNM-based PSD estimator results in a significantly better performance than the ML-based PSD estimator and a similar or slightly better performance than the EVD-based PSD estimator when used in an MWF with erroneously estimated RETFs.

5. CONCLUSION

In this paper a multi-channel late reverberant PSD estimator based on nuclear norm minimization has been proposed, which does not require an estimate of the RETFs. In order to account for modeling or estimation errors in the estimated reverberant PSD matrix, this matrix is modeled as the sum of a low rank matrix and a scaled diffuse coherence matrix. Among all pairs of scalars and matrices which yield feasible decompositions, the late reverberation PSD is estimated as the scalar associated with the matrix of minimum rank. Instead of minimizing the non-convex matrix rank, it has been proposed to use a convex relaxation approach and estimate the late reverberation PSD by minimizing the nuclear norm. Experimental results have shown that using the proposed NNM-based PSD estimator in an MWF for speech dereverberation yields a similar or better performance than the ML-based and EVD-based estimators.

6. REFERENCES

[1] J. S. Bradley, H. Sato, and M. Picard, "On the importance of early reflections for speech in rooms," *Journal of the Acoustical Society of America*, vol. 113, no. 6, pp. 3233–3244, June 2003.

[2] 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 algorithms," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sept. 2014, pp. 333–337.

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

[4] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. dissertation, Technische Universiteit Eindhoven, Eindhoven, Netherlands, June 2007.

[5] 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.

[6] B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukić, T. Gerkmann, S. Doclo, and S. Goetze, "Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, 2015.

[7] K. Lebart and J. M. Boucher, "A new method based on spectral subtraction for speech dereverberation," *Acta Acoustica*, vol. 87, no. 3, pp. 359–366, May-Jun. 2001.

[8] E. A. P. Habets, S. Gannot, and I. Cohen, "Late reverberant spectral variance estimation based on a statistical model," *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–774, Sept. 2009.

[9] 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.

[10] 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, pp. 61–65.

[11] 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 *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, Oct. 2015.

[12] 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.

[13] 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.

[14] 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.

[15] O. Schwartz, S. Gannot, and E. A. P. Habets, "Joint estimation of late reverberant and speech power spectral densities in noisy environments using Frobenius norm," in *Proc. European Signal Processing Conference*, Budapest, Hungary, Sept. 2016, pp. 1123–1127.

[16] I. Kodrasi and S. Doclo, "EVD-based multi-channel dereverberation of a moving speaker using different RETF estimation methods," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, San Francisco, USA, Mar. 2017, pp. 116–120.

[17] ——, "Late reverberant power spectral density estimation based on an eigenvalue decomposition," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, New Orleans, USA, Mar. 2017, pp. 611–615.

[18] M. Fazel, "Matrix rank minimization with applications," Ph.D. dissertation, Stanford University, California, USA, Mar. 2002.

[19] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the Association for Computing Machinery*, vol. 58, no. 3, pp. 11:1–11:37, June 2011.

[20] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[21] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson Jr., "Measurement of correlation coefficients in reverberant sound fields," *Journal of the Acoustical Society of America*, vol. 27, no. 6, pp. 1072–1077, Nov. 1955.

[22] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, Aug. 2010.

[23] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[24] 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.

[25] 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.

[26] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "The ACE challenge - Corpus description and performance evaluation," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, Oct. 2015.

[27] 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.

[28] 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.

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