Cramér-Rao Bound Analysis of Reverberation Level Estimators for Dereverberation and Noise Reduction

Ofer Schwartz, *Student Member, IEEE,* Sharon Gannot, *Senior Member, IEEE,* and Emanuël A. P. Habets, *Senior Member, IEEE*

***Abstract*—The reverberation power spectral density (PSD) is often required for dereverberation and noise reduction algorithms. In this work, we compare two maximum likelihood (ML) estimators of the reverberation PSD in a noisy environment. In the first estimator, the direct path is first blocked. Then, the ML criterion for estimating the reverberation PSD is stated according to the probability density function (p.d.f.) of the blocking matrix (BM) outputs. In the second estimator, the speech component is not blocked. Since the anechoic speech PSD is usually unknown in advance, it is estimated as well. To compare the expected mean square error (MSE) between the two ML estimators of the reverberation PSD, the Cramér-Rao Bounds (CRBs) for the two ML estimators are derived. We show that the CRB for the joint reverberation and speech PSD estimator is lower than the CRB for estimating the reverberation PSD from the BM outputs. Experimental results show that the MSE of the two estimators indeed obeys the CRB curves. Experimental results of multi-microphone dereverberation and noise reduction algorithm show the benefits of using the ML estimators in comparison with another baseline estimators.**

I. Introduction

Reverberation and ambient noise may degrade the quality and intelligibility of the signals captured by the microphones of mobile devices, smart TVs and audio conferencing systems. While intelligibility does not degrade by early reflections, it can be significantly degraded by late reflections (i.e., reverberation) due to overlap masking effects [1]. The estimation of the reverberation power spectral density (PSD) is a much more challenging task than the estimation of the ambient noise PSD, since it is highly non-stationary and since speech-absence periods cannot be utilized.

Both single- and multi-microphone techniques have been proposed to reduce reverberation (see [2], [3] and the references therein), while many of these techniques require an estimate of the reverberation PSD (e.g. [4]). A statistical model of reverberant signals usually depends on the PSD of the anechoic speech component. Since the anechoic speech PSD is changing rapidly across time and is unknown in advance, the anechoic speech PSD should be either ignored or estimated as well. In our previous work [4], a multi-microphone minimum mean square error (MMSE) estimator of the early speech component was implemented as a minimum variance distortionless response (MVDR) beamformer followed by a postfilter. The reverberation and the ambient noise were treated by both the MVDR stage and the postfiltering stage. The ambient noise PSD matrix was assumed to be known. The most difficult task was to provide an accurate estimate of the reverberation PSD that is required for both stages. The reverberation PSD was estimated by averaging the marginal reverberation levels at the microphones, obtained using the single-channel estimator proposed in [5]. The postfilter was calculated using the decision-directed approach [6]. The speech PSD was not estimated explicitly. In the sequel, multichannel estimators of the speech, noise or reverberation PSDs are discussed.

In [7], the authors proposed an estimator for the noise PSD matrix assuming a non-reverberant scenario and time-varying noise level. A normalized version of the noise PSD matrix was estimated, using the most recent speech-absent segments. Next, for the speech-present segments, only the level of the noise was estimated. The authors proposed to estimate the level of the noise from signals at the output of a blocking matrix (BM), which blocks the speech signals. Since the speech signal is blocked, the PSD matrix of the BM outputs comprises only noise components. A closed-form maximum likelihood (ML) estimator of the noise level was proposed w.r.t. the probability density function (p.d.f.) of the BM outputs. This estimator may be applied also for reverberation assuming that the observed signals do not contain additive noise.

In [8], [9], a reverberant and noisy environment was addressed. The reverberation was modelled as a diffuse sound field with time-varying level and the noise PSD

O. Schwartz and S. Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan 5290002, Israel (e-mail: ofer.shwartz@live.biu.ac.il; sharon.gannot@biu.ac.il).

E. A. P. Habets is with the International Audio Laboratories Erlangen (a joint institution between the University of Erlangen-Nuremberg and Fraunhofer IIS), 91058 Erlangen, Germany (e-mail: emanuel.habets@audiolabserlangen.de).

matrix was assumed to be known. In this work, the authors proposed to estimate the time-varying level of the reverberation from the signals at the output of a BM. The authors defined an error matrix using the PSD matrix of the BM outputs. The elements of the error matrix were assumed to be zero-mean Gaussian variables. The best fitting value for the reverberation PSD was estimated by minimizing the Frobenius norm of the error matrix.

Recently, in [10], an estimator for the reverberation PSD in a noisy environment was proposed. First, the received signals are filtered by a BM to block the anechoic speech. Then, the likelihood of the reverberation PSD, given the signals at the output of the BM, is maximized. Due to the complexity of the p.d.f., a closed-form solution cannot be derived. Instead, an iterative Newton's method for maximizing the ML was used. For the application of the Newton's iterations, the first- and second-order derivatives of the log-likelihood are calculated in closed-form. However, since the BM processes the data and reduces the number of available signals, applying the ML at the output of the BM is prone to sub-optimality.

In [11], the dereverberation task for hearing aids applications was addressed, assuming a noise-free environment. The reverberation component was similarly modeled as in [8]. The authors propose not to block the speech signal but to rather estimate its PSD as well. A closed-form solution for the maximum likelihood estimators (MLEs) of the time-varying reverberation PSD and the anechoic speech PSD is then presented based on an expression derived in [12]. The MLE was derived under the assumption that the speech component and the reverberant component are zero-mean Gaussian vectors, with rank-1 PSD matrix for the speech component and full-rank PSD matrix for the reverberant component.

More recently, in [13], an optimal estimator in the ML sense for the reverberation PSD in noisy environment was proposed without using a blocking stage. Instead, the reverberation and the anechoic speech PSD levels were jointly estimated. Similarly to [10], a closed-form solution cannot be derived. Thus, an iterative Newton's method for maximizing the ML was used. For the application of Newton's iterations, the first- and second-order derivatives of the log-likelihood were calculated in closed-form. It is unclear whether the preliminarily blocking stage used in [10] degrades or enhances the accuracy of the reverberation estimation. It is therefore interesting to compare the two MLEs, i.e. the MLE of the reverberant PSD without using the BM [13] and the MLE which uses the BM [10].

A common way to analyze an MLE is by examining its Cramér-Rao Bound (CRB), since the mean square error (MSE) of an MLE asymptotically achieves the CRB. By careful examination of the CRB expression, one can reveal factors that may decrease the MSE of the MLE. In [14], the authors derive the CRB expressions for the MLE of the reverberant and speech PSDs proposed in [11], [12]. The authors show that the CRB expression for the reverberation PSD is independent of the speech component. It should be emphasized that this result is valid only for the noiseless case. In [15], the authors show that the MLE derived in [11], which circumvents the blocking of the desired speech, has a lower MSE than the MLE derived in [8], which uses the BM.

In the current paper, the CRB for the estimator in [10] that uses the blocking stage, and the CRB for the joint estimator in [13] that circumvents the blocking stage are derived and compared. Although the estimators are based on different signal models, they both provide an exact solution in the ML sense given the considered signal model. The ML estimator is commonly used in the signal processing community as it is mathematically rigorous. Moreover, the CRB is a reliable tool for checking the estimation error of the ML estimator. The comparison between these two estimators is interesting from both practical and theoretical points of view as it provides new insights into the pros and cons of using a blocking stage, which is used in various noise reduction and dereverberation algorithms (see references [7]–[10]).

The paper is organized as follows. In Section II, the reverberant and noisy signal model is formulated. In Section III, the two ML estimators are derived and summarized. In Section IV, the two CRB expressions are derived and compared. Section V presents the simulation study by calculating the MSE of the two MLEs and their respective CRB and demonstrates the dereverberation and noise reduction capabilities of an algorithm that uses these estimators. Finally, conclusions are drawn in Section VI and the work is summarized.

II. Problem Formulation

We consider N microphone observations consisting of reverberant speech and additive noise. The reverberant speech can be split into two components, the direct and reverberation parts:

$$Y_i(m,k) = X_i(m,k) + R_i(m,k) + V_i(m,k), \qquad (1)$$

where $Y_i(m,k)$ denotes the ith microphone observation in time-index m and frequency bin index k, $X_i(m,k)$ denotes the direct speech component, $R_i(m,k)$ denotes the reverberation and $V_i(m,k)$ denotes the ambient noise. Here $X_i(m,k)$ is modeled as a multiplication between

the anechoic speech and the direct transfer function (DTF)

$$X_i(m,k) = G_{\mathrm{d},i}(k)S(m,k), \tag{2}$$

where $G_{\mathrm{d},i}(k)$ is the DTF and $S(m,k)$ is the desired anechoic speech.

The N microphone signals are concatenated in the vector

$$\mathbf{y}(m,k) = \mathbf{x}(m,k) + \mathbf{r}(m,k) + \mathbf{v}(m,k) \tag{3}$$

where

$$\begin{aligned}
\mathbf{y}(m,k) &= \left[\begin{array}{cccc} Y_1(m,k) & Y_2(m,k) & \dots & Y_N(m,k) \end{array}\right]^{\mathrm{T}} \\
\mathbf{x}(m,k) &= \left[\begin{array}{cccc} X_1(m,k) & X_2(m,k) & \dots & X_N(m,k) \end{array}\right]^{\mathrm{T}} \\
&= \mathbf{g}_{\mathrm{d}}(k)S(m,k) \\
\mathbf{g}_{\mathrm{d}}(k) &= \left[\begin{array}{cccc} G_{\mathrm{d},1}(k) & G_{\mathrm{d},2}(k) & \dots & G_{\mathrm{d},N}(k) \end{array}\right]^{\mathrm{T}} \\
\mathbf{r}(m,k) &= \left[\begin{array}{cccc} R_1(m,k) & R_2(m,k) & \dots & R_N(m,k) \end{array}\right]^{\mathrm{T}} \\
\mathbf{v}(m,k) &= \left[\begin{array}{cccc} V_1(m,k) & V_2(m,k) & \dots & V_N(m,k) \end{array}\right]^{\mathrm{T}}.
\end{aligned}$$

The speech signal is modeled as a Gaussian process with $S(m,k) \sim \mathcal{N}(0, \phi_S(m,k))$ where $\phi_S(m,k)$ is the speech PSD. The reverberation and the noise components are assumed to be uncorrelated and may be modelled by zero-mean multivariate Gaussian probability. The PSD matrix of the noise $\mathbf{\Phi}_{\mathbf{v}}(k) = E\left\{\mathbf{v}(m,k)\mathbf{v}^{\mathrm{H}}(m,k)\right\}$ is assumed to be time-invariant and known in advance or can be accurately estimated during speech-absent periods. The PSD matrix of the reverberation is time-varying, since the reverberation originates from the speech source. However, the spatial characteristic of the reverberation is assumed to be time-invariant, as long as the speaker and microphone positions do not change. Therefore, it is reasonable to model the PSD matrix of the reverberation as a time-invariant matrix with time-varying levels. Hence, the p.d.f. of the reverberant signal vector is modelled as:

$$\mathbf{r}(m,k) \sim \mathcal{N}\left(0,\, \phi_R(m,k)\,\mathbf{\Gamma}(k)\right), \tag{4}$$

where $\mathbf{\Gamma}(k)$ is the time-invariant spatial coherence matrix of the reverberation and $\phi_R(m,k)$ is the temporal level of the reverberation. In the current work, we assume that the reverberation can be modelled using a spatially homogenous and spherically isotropic sound field (as in many previous works [4], [16]–[19]). Under this assumption, and assuming omnidirectional microphones and a known microphone array geometry, the spatial coherence is uniquely defined as

$$\mathbf{\Gamma}_{ij}(k) = \mathrm{sinc}\left(\frac{2\pi k}{K}\frac{\mathcal{D}_{i,j}}{T_{\mathrm{s}}c}\right), \tag{5}$$

where $\mathrm{sinc}(x) = \sin(x)/x$, K is the number of frequency bins, $\mathcal{D}_{i,j}$ is the inter-distance between microphones i and j, T_{s} is the sampling time and c is the sound velocity. The goal in this work is to compare two ML estimators of $\phi_R(m,k)$, given the measurements $\mathbf{y}(m,k)$, with the other parameters $\mathbf{g}_{\mathrm{d}}(k)$, $\mathbf{\Gamma}(k)$, and $\mathbf{\Phi}_{\mathbf{v}}(k)$ known. As the speech PSD $\phi_S(m,k)$ is naturally unavailable, $\phi_S(m,k)$ should be estimated or eliminated. One alternative is to simultaneously estimate the variance of the reverberation and the speech. Another alternative is to first block the speech signal using a standard BM, and then estimate the reverberation variance alone. The main goal of the paper is to analytically compare two ML estimators of the reverberation power. The main contributions of the paper are 1) to describe the ML estimators with and without using the blocking matrix; 2) to derivate the CRBs of each estimator; 3) to analytical compare these CRBs; 4) to discuss the considerations for choosing an appropriate estimator for a given scenario; and in particular 5) to analytically study the influence of applying the blocking stage on the estimation error. This result may also be applicable to other algorithms that use a blocking matrix.

In the following sections we first derive the MLE (Section III) and then derive and compare their respective CRBs (Section IV).

III. Presentation of the Two Maximum Likelihood Schemes

First, the joint MLE of both the reverberation and the anechoic speech PSDs is derived. Second, the MLE is derived for the reverberant PSD alone, while the speech signal is blocked. In the following, whenever applicable, the frequency index k and the time index m are omitted for brevity.

A. Joint estimation of reverberation and speech PSDs

The PSD is foreseen to be smooth over time. It is therefore proposed to estimate ϕ_S and ϕ_R using a set of successive segments by concatenating several snapshots, assumed to be i.i.d. Denote $\bar{\mathbf{y}}$ as a vector of M i.i.d. past snapshots of $\mathbf{y}$,

$$\bar{\mathbf{y}}(m) \equiv \left[\begin{array}{ccc} \mathbf{y}^{\mathrm{T}}(m-M+1) & \dots & \mathbf{y}^{\mathrm{T}}(m) \end{array}\right]^{\mathrm{T}}. \tag{6}$$

Collecting all definitions, the p.d.f. of microphone signals $\bar{\mathbf{y}}$ is

$$\begin{aligned} f\left(\bar{\mathbf{y}}(m)\right) &= \prod_{\tilde{m}=m-M+1}^{m} \frac{1}{\pi^N|\mathbf{\Phi_y}|}\exp\left(\mathbf{y}^{\mathrm{H}}(\tilde{m})\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{y}(\tilde{m})\right) \\ &= \left(\frac{1}{\pi^N|\mathbf{\Phi_y}|}\right)^M \exp\left(\sum_{\tilde{m}=m-M+1}^{m}\mathbf{y}^{\mathrm{H}}(\tilde{m})\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{y}(\tilde{m})\right) \\ &= \left(\frac{1}{\pi^N|\mathbf{\Phi_y}|}\right)^M \exp\left(\mathrm{Tr}\left[\mathbf{\Phi}_{\mathbf{y}}^{-1}\ M\ \mathbf{R}_{\bar{\mathbf{y}}}(m)\right]\right) \\ &= \left(\frac{1}{\pi^N|\mathbf{\Phi_y}|}\exp\left(\mathrm{Tr}\left[\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{R}_{\bar{\mathbf{y}}}(m)\right]\right)\right)^M, \end{aligned} \tag{7}$$

where $\mathbf{\Phi_y}$ is the PSD matrix of the observed signals that is assumed to be fixed during the entire segment and is defined as

$$\mathbf{\Phi_y} = \phi_S \mathbf{g}_{\mathrm{d}}\mathbf{g}_{\mathrm{d}}^{\mathrm{H}} + \phi_R\,\mathbf{\Gamma} + \mathbf{\Phi_v} \tag{8}$$

and

$$\mathbf{R}_{\bar{\mathbf{y}}}(m) \equiv \frac{1}{M}\sum_{\tilde{m}=m-M+1}^{m}\mathbf{y}(\tilde{m})\mathbf{y}^{\mathrm{H}}(\tilde{m}). \tag{9}$$

The MLE aims at finding the most likely parameter set $\boldsymbol{\phi} = \left[\ \phi_R \quad \phi_S\ \right]^{\mathrm{T}}$ w.r.t. the p.d.f. of $\bar{\mathbf{y}}(m)$, i.e.,

$$\boldsymbol{\phi}^{\mathrm{ML},\bar{\mathbf{y}}} = \underset{\boldsymbol{\phi}}{\mathrm{argmax}}\log f\left(\bar{\mathbf{y}}(m);\,\boldsymbol{\phi}\right). \tag{10}$$

Since there is no closed-form solution to $\boldsymbol{\phi}^{\mathrm{ML},\bar{\mathbf{y}}}$, the Newton's method may be applied to derive an iterative search (c.f. [20])

$$\boldsymbol{\phi}^{(\ell+1)} = \boldsymbol{\phi}^{(\ell)} - \mathbf{H}^{-1}\left(\boldsymbol{\phi}^{(\ell)}\right)\mathbf{d}\left(\boldsymbol{\phi}^{(\ell)}\right), \tag{11}$$

where $\mathbf{d}\left(\boldsymbol{\phi}\right)$ is the first-order derivative of the log-likelihood with respect to $\boldsymbol{\phi}$, and $\mathbf{H}\left(\boldsymbol{\phi}\right)$ is the corresponding Hessian matrix, i.e.,

$$\begin{aligned} \mathbf{d}\left(\boldsymbol{\phi}\right) &\equiv \frac{\partial \log f\left(\bar{\mathbf{y}}(m);\,\boldsymbol{\phi}\right)}{\partial \boldsymbol{\phi}}, \\ \mathbf{H}\left(\boldsymbol{\phi}\right) &\equiv \frac{\partial^2 \log f\left(\bar{\mathbf{y}}(m);\,\boldsymbol{\phi}\right)}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}^{\mathrm{T}}}. \end{aligned} \tag{12}$$

The first-order derivative $\mathbf{d}\left(\boldsymbol{\phi}\right)$ is a 2-dimensional vector

$$\mathbf{d}\left(\boldsymbol{\phi}\right) \equiv \left[\ D_R\left(\boldsymbol{\phi}\right) \quad D_S\left(\boldsymbol{\phi}\right)\ \right]^{\mathrm{T}}, \tag{13}$$

with elements

$$D_i\left(\boldsymbol{\phi}\right) = M\ \mathrm{Tr}\left[\left(\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{R}_{\bar{\mathbf{y}}}(m) - \mathbf{I}\right)\mathbf{\Phi}_{\mathbf{y}}^{-1}\frac{\partial \mathbf{\Phi_y}}{\partial \phi_i}\right], \tag{14}$$

where $i \in \{R, S\}$, $\frac{\partial \mathbf{\Phi_y}(m)}{\partial \phi_R} = \mathbf{\Gamma}$, $\frac{\partial \mathbf{\Phi_y}}{\partial \phi_S} = \mathbf{g}_{\mathrm{d}}\mathbf{g}_{\mathrm{d}}^{\mathrm{H}}$. The Hessian is a 2×2 matrix:

$$\mathbf{H}\left(\boldsymbol{\phi}\right) \equiv \left[\begin{array}{cc} H_{RR}\left(\boldsymbol{\phi}\right) & H_{SR}\left(\boldsymbol{\phi}\right) \\ H_{RS}\left(\boldsymbol{\phi}\right) & H_{SS}\left(\boldsymbol{\phi}\right) \end{array}\right]. \tag{15}$$

Algorithm 1: Multi-microphone reverberation and speech PSD estimation in a noisy environment.

for *all time frames and frequency bins* m, k **do**
  Compute $\mathbf{R}_{\bar{\mathbf{y}}}(m)$ using (9).
  Initialize by $\boldsymbol{\phi}^{(0)} = \mathbf{0}$.
  if $\left(\mathbf{d}_1\left(\mathbf{0}\right) < 0\right)\,\&\,\left(\mathbf{d}_2\left(\mathbf{0}\right) < 0\right)$ **then**
    $\phi_R^{\mathrm{ML},\bar{\mathbf{y}}} = \phi_S^{\mathrm{ML},\bar{\mathbf{y}}} = \epsilon$
  else
    for $\ell = 0$ **to** $L-1$ **do**
      Compute $\mathbf{\Phi}_{\mathbf{y}}^{(\ell)}$ using (8)
      Calculate $\boldsymbol{\phi}^{(\ell+1)}$ using (11).
      Bound $\boldsymbol{\phi}^{(\ell+1)}$ to the range $[\epsilon, Z]$.
    end
  end
  $\boldsymbol{\phi}^{\mathrm{ML},\bar{\mathbf{y}}} = \boldsymbol{\phi}^{(L)}$
end

Applying a second derivative to $D_R\left(\boldsymbol{\phi}\right)$ and $D_S\left(\boldsymbol{\phi}\right)$ yields the elements of $\mathbf{H}\left(\boldsymbol{\phi}\right)$:

$$\begin{aligned} H_{ij}\left(\boldsymbol{\phi}\right) = -M\ \mathrm{Tr}\Bigg[&\mathbf{\Phi}_{\mathbf{y}}^{-1}\frac{\partial \mathbf{\Phi_y}}{\partial \phi_j}\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{R}_{\bar{\mathbf{y}}}(m)\mathbf{\Phi}_{\mathbf{y}}^{-1}\frac{\partial \mathbf{\Phi_y}}{\partial \phi_i} + \\ &\left(\mathbf{\Phi}_{\mathbf{y}}^{-1}\mathbf{R}_{\bar{\mathbf{y}}}(m) - \mathbf{I}\right)\mathbf{\Phi}_{\mathbf{y}}^{-1}\frac{\partial \mathbf{\Phi_y}}{\partial \phi_j}\mathbf{\Phi}_{\mathbf{y}}^{-1}\frac{\partial \mathbf{\Phi_y}}{\partial \phi_i}\Bigg], \end{aligned} \tag{16}$$

where $i, j \in \{R, S\}$.

In order to apply the Newton's method, the following practical considerations has to been taken into account:

1) The PSDs of the reverberation and the speech are always positive. Thus, the Newton's method may be initialized with $\boldsymbol{\phi}^{(0)} = \mathbf{0}$.
2) In cases where the first-order derivatives $\mathbf{d}\left(\boldsymbol{\phi}\right)$ in $\boldsymbol{\phi} = [0, 0]^{\mathrm{T}}$ are negative, the maximum point is probably located in the third quadrant (i.e., $\phi_S < 0$ and $\phi_R < 0$). In such a case, the ML estimates of $\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}$ and $\phi_S^{\mathrm{ML},\bar{\mathbf{y}}}$ can be set to be zero without executing the Newton's method. However, to avoid dividing by zero, the reverberation and speech levels can be set to some small positive value ϵ.
3) To avoid illegal PSD level estimates, the parameter set $\boldsymbol{\phi}^{(\ell)}$ may be confined to be in the range $[\epsilon, Z]$ where $Z \equiv \frac{1}{N}\mathbf{y}^{\mathrm{H}}\mathbf{y} - \frac{1}{N}\mathrm{Tr}\left[\mathbf{\Phi_v}(k)\right]$ is the a posteriori level of the entire reverberant speech component.

The joint ML estimation of $\boldsymbol{\phi}$ is summarized in Algorithm 1. Additional technical details can be found in [13].

For a noiseless case (which can be determined by evaluating $\mathrm{Tr}\left[\mathbf{\Phi_v}\right]$), a closed-form solution for ϕ_R is

available [11, Eq. 7a]:

$$\phi_R^{\mathrm{ML},\bar{\mathbf{y}}} = \frac{1}{N-1}\mathrm{Tr}\left[\left(\mathbf{\Gamma}^{-1} - \frac{\mathbf{\Gamma}^{-1}\mathbf{g}_\mathrm{d}\mathbf{g}_\mathrm{d}^\mathrm{H}\mathbf{\Gamma}^{-1}}{\mathbf{g}_\mathrm{d}^\mathrm{H}\mathbf{\Gamma}^{-1}\mathbf{g}_\mathrm{d}}\right)\mathbf{R}_{\bar{\mathbf{y}}}(m)\right]. \tag{17}$$

B. Estimation of the reverberation PSD using a BM

In this section, the speech component is first blocked, and then an MLE of the reverberation PSD is derived. The speech signal is blocked using a standard blocking matrix, satisfying $\mathbf{B}^\mathrm{H}\mathbf{g}_\mathrm{d} = 0$. A sparse option for the blocking matrix is (see e.g. [4], [21])

$$\mathbf{B} = \begin{bmatrix} -\frac{G_{\mathrm{d},2}^*}{G_{\mathrm{d},1}^*} & -\frac{G_{\mathrm{d},3}^*}{G_{\mathrm{d},1}^*} & \dots & -\frac{G_{\mathrm{d},N}^*}{G_{\mathrm{d},1}^*} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{18}$$

The output of the blocking matrix is given by

$$\mathbf{u} = \mathbf{B}^\mathrm{H}\mathbf{y} = \mathbf{B}^\mathrm{H}(\mathbf{r}+\mathbf{v}). \tag{19}$$

The p.d.f. of a set of M i.i.d. snapshots $\bar{\mathbf{u}}(m)$ (where $\bar{\mathbf{u}}(m)$ is similarly defined as (6)) is given by

$$f(\bar{\mathbf{u}}(m)) = \left(\frac{1}{\pi^N|\mathbf{\Phi}_\mathbf{u}|}\exp\left(\mathrm{Tr}\left[\mathbf{\Phi}_\mathbf{u}^{-1}\mathbf{R}_{\bar{\mathbf{u}}}(m)\right]\right)\right)^M, \tag{20}$$

where

$$\mathbf{\Phi}_\mathbf{u}(m) = \mathbf{B}^\mathrm{H}\left(\phi_R\mathbf{\Gamma} + \mathbf{\Phi}_V\right)\mathbf{B} \tag{21}$$

and

$$\mathbf{R}_{\bar{\mathbf{u}}}(m) = \frac{1}{M}\sum_{\tilde{m}=m-M+1}^{m}\mathbf{u}(\tilde{m})\mathbf{u}^\mathrm{H}(\tilde{m}) = \mathbf{B}^\mathrm{H}\mathbf{R}_{\bar{\mathbf{y}}}(m)\mathbf{B}. \tag{22}$$

Under this model, the MLE of ϕ_R aims at finding the most likely ϕ_R w.r.t. the p.d.f. of $\bar{\mathbf{u}}(m)$, i. e.,

$$\phi_R^{\mathrm{ML},\bar{\mathbf{u}}} = \underset{\phi_R}{\mathrm{argmax}}\log f\left(\bar{\mathbf{u}}(m);\,\phi_R\right). \tag{23}$$

Similar to the previous estimator, there is no closed-form solution for $\phi_R^{\mathrm{ML},\bar{\mathbf{u}}}$. Newton's method may be used to derive an iterative search (c.f. [20])

$$\phi_R^{(\ell+1)} = \phi_R^{(\ell)} - \frac{\mathcal{D}\left(\phi_R^{(\ell)}\right)}{\mathcal{H}\left(\phi_R^{(\ell)}\right)}, \tag{24}$$

where $\mathcal{D}(\phi_R)$ and $\mathcal{H}(\phi_R)$ are the first- and second-order derivatives

$$\mathcal{D}(\phi_R) \equiv \frac{\partial\log f\left(\bar{\mathbf{u}}(m);\,\phi_R\right)}{\partial\phi_R}, \tag{25}$$

$$\mathcal{H}(\phi_R) \equiv \frac{\partial^2\log f\left(\bar{\mathbf{u}}(m);\,\phi_R\right)}{\partial\phi_R^2}. \tag{26}$$

Algorithm 2: Multi-microphone reverberation PSD estimation.

```
for all time frames and frequency bins m, k do
    Compute R_ū(m) using (22).
    if D(0) < 0 then
        φ_R^{ML,ū} = ε
    else
        Initialize by φ_R^{(0)} = 0.
        for ℓ = 0 to L − 1 do
            Compute Φ_u^{(ℓ)} using (21)
            Calculate φ_R^{(ℓ+1)} using (24).
            Bound φ_R^{(ℓ+1)} to the range [ε, Z].
        end
        φ_R^{ML,ū} = φ_R^{(L)}
    end
end
```

The first-order derivative equals to

$$\mathcal{D}(\phi_R) = M\,\mathrm{Tr}\left[\left(\mathbf{\Phi}_\mathbf{u}^{-1}\mathbf{R}_{\bar{\mathbf{u}}}(m) - \mathbf{I}\right)\mathbf{\Phi}_\mathbf{u}^{-1}\frac{\partial\mathbf{\Phi}_\mathbf{u}}{\partial\phi_R}\right], \tag{27}$$

where $\frac{\partial\mathbf{\Phi}_\mathbf{u}}{\partial\phi_R} = \mathbf{B}^\mathrm{H}\mathbf{\Gamma}\mathbf{B}$. The second-order derivative equals to

$$\mathcal{H}(\phi_R) = -M\,\mathrm{Tr}\left[\mathbf{\Phi}_\mathbf{u}^{-1}\frac{\partial\mathbf{\Phi}_\mathbf{u}}{\partial\phi_R}\mathbf{\Phi}_\mathbf{u}^{-1}\mathbf{R}_{\bar{\mathbf{u}}}(m)\mathbf{\Phi}_\mathbf{u}^{-1}\frac{\partial\mathbf{\Phi}_\mathbf{u}}{\partial\phi_R} + \left(\mathbf{\Phi}_\mathbf{u}^{-1}\mathbf{R}_{\bar{\mathbf{u}}}(m) - \mathbf{I}\right)\mathbf{\Phi}_\mathbf{u}^{-1}\frac{\partial\mathbf{\Phi}_\mathbf{u}}{\partial\phi_R}\mathbf{\Phi}_\mathbf{u}^{-1}\frac{\partial\mathbf{\Phi}_\mathbf{u}}{\partial\phi_R}\right]. \tag{28}$$

Similar practical considerations to the described in the previous section can be applied to the estimator which uses the BM. The MLE of ϕ_R which uses the BM is summarized in Algorithm 2. Additional technical details can be found in [10].

For the noiseless case, a closed-form solution for ϕ_R is available [7, Eq. 7]:

$$\phi_R^{\mathrm{ML},\bar{\mathbf{u}}} = \frac{1}{N-1}\mathrm{Tr}\left[\left(\mathbf{B}^\mathrm{H}\mathbf{\Gamma}\mathbf{B}\right)^{-1}\mathbf{R}_{\bar{\mathbf{u}}}(m)\right]. \tag{29}$$

Interestingly, the estimators in (17) and in (29) are mathematically identical. In Appendix A, the following identity is proved

$$\mathbf{\Gamma}^{-1} - \frac{\mathbf{\Gamma}^{-1}\mathbf{g}_\mathrm{d}\mathbf{g}_\mathrm{d}^\mathrm{H}\mathbf{\Gamma}^{-1}}{\mathbf{g}_\mathrm{d}^\mathrm{H}\mathbf{\Gamma}^{-1}\mathbf{g}_\mathrm{d}} = \mathbf{B}\left(\mathbf{B}^\mathrm{H}\mathbf{\Gamma}\mathbf{B}\right)^{-1}\mathbf{B}^\mathrm{H}. \tag{30}$$

Substituting the above identity in (17) and using the cyclic property of the trace operation on $\mathbf{B}$ yields the estimator in (29). Hence, in the absence of noise the two ML estimators are identical.

IV. CRB ANALYSIS OF THE ML ESTIMATORS

In the following section, the CRBs of the estimators summarized in Section III are derived and compared. By comparing the CRBs, some insights about the estimators may be deduced.

A. CRB Derivation for the Joint Estimator

In the following section, the CRB for ϕ_R under the joint estimator of the reverberation and speech PSDs is derived. To derive the CRB, the Fisher information matrix (FIM) of $\boldsymbol{\phi}^{\mathrm{ML},\bar{\mathbf{y}}}$ is first derived. The FIM of ϕ_R and ϕ_S w.r.t. the p.d.f. of $\bar{\mathbf{y}}$ can be decomposed as

$$\mathbf{I}^{\bar{\mathbf{y}}} = \begin{pmatrix} \mathrm{I}^{\bar{\mathbf{y}}}_{RR} & \mathrm{I}^{\bar{\mathbf{y}}}_{SR} \\ \mathrm{I}^{\bar{\mathbf{y}}}_{RS} & \mathrm{I}^{\bar{\mathbf{y}}}_{SS} \end{pmatrix}, \tag{31}$$

where $\mathrm{I}^{\bar{\mathbf{y}}}_{ij}$ is the Fisher information of ϕ_i and ϕ_j (where $i,j \in \{R,S\}$). The CRB for ϕ_R is obtained by inverting the FIM and taking the corresponding component:

$$\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R} = \left\{ \left(\mathbf{I}^{\bar{\mathbf{y}}}\right)^{-1} \right\}_{11} = \frac{\mathrm{I}^{\bar{\mathbf{y}}}_{SS}}{\mathrm{I}^{\bar{\mathbf{y}}}_{RR}\mathrm{I}^{\bar{\mathbf{y}}}_{SS} - \mathrm{I}^{\bar{\mathbf{y}}}_{RS}\mathrm{I}^{\bar{\mathbf{y}}}_{SR}}. \tag{32}$$

Both ϕ_R and ϕ_S are components of the PSD matrix $\boldsymbol{\Phi}_{\mathbf{y}}$. The Fisher information of two parameters associated with a PSD matrix of a Gaussian vector is given by [22], [12] , [23]:

$$\mathrm{I}^{\bar{\mathbf{y}}}_{ij} = M \ \mathrm{Tr}\left[\boldsymbol{\Phi}_{\mathbf{y}}^{-1} \frac{\partial \boldsymbol{\Phi}_{\mathbf{y}}}{\partial \phi_i} \boldsymbol{\Phi}_{\mathbf{y}}^{-1} \frac{\partial \boldsymbol{\Phi}_{\mathbf{y}}}{\partial \phi_j} \right], \tag{33}$$

where M is the number of snapshots. Substituting the derivatives $\frac{\partial \boldsymbol{\Phi}_y}{\partial \phi_R}$ and $\frac{\partial \boldsymbol{\Phi}_{\mathbf{y}}}{\partial \phi_S}$ in (33) yields the expressions for the elements of the FIM:

$$\mathrm{I}^{\bar{\mathbf{y}}}_{RR} = M \ \mathrm{Tr}\left[\boldsymbol{\Phi}_{\mathbf{y}}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Phi}_{\mathbf{y}}^{-1} \boldsymbol{\Gamma} \right], \tag{34a}$$

$$\mathrm{I}^{\bar{\mathbf{y}}}_{SS} = M \ \mathrm{Tr}\left[\boldsymbol{\Phi}_{\mathbf{y}}^{-1} \mathbf{g}_{\mathrm{d}} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Phi}_{\mathbf{y}}^{-1} \mathbf{g}_{\mathrm{d}} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \right], \tag{34b}$$

$$\mathrm{I}^{\bar{\mathbf{y}}}_{RS} = I^{\bar{\mathbf{y}}}_{SR} = M \ \mathrm{Tr}\left[\boldsymbol{\Phi}_{\mathbf{y}}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Phi}_{\mathbf{y}}^{-1} \mathbf{g}_{\mathrm{d}} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \right]. \tag{34c}$$

To continue the derivation, $\boldsymbol{\Phi}_{\mathbf{y}}^{-1}$ is calculated using the matrix-inversion Lemma:

$$\boldsymbol{\Phi}_{\mathbf{y}}^{-1} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1} \mathbf{g}_{\mathrm{d}} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Psi}^{-1} \frac{1}{\phi_S^{-1} + \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Psi}^{-1} \mathbf{g}_{\mathrm{d}}}, \tag{35}$$

where

$$\boldsymbol{\Psi} \equiv \phi_R \boldsymbol{\Gamma} + \boldsymbol{\Phi}_{\mathbf{v}}. \tag{36}$$

Substituting (35) in (34) and employing some algebraic steps, the following results are obtained:

$$\mathrm{I}^{\bar{\mathbf{y}}}_{RR} = M \left(\gamma_0 - \frac{2\gamma_3}{\gamma_1 + \phi_S^{-1}} + \frac{\gamma_2^2}{\left(\gamma_1 + \phi_S^{-1}\right)^2} \right) \tag{37a}$$

$$\mathrm{I}^{\bar{\mathbf{y}}}_{SS} = \frac{M\gamma_1^2}{\phi_S^2 \left(\gamma_1 + \phi_S^{-1}\right)^2} \tag{37b}$$

$$\mathrm{I}^{\bar{\mathbf{y}}}_{SR} = \frac{M\gamma_2}{\phi_S^2 \left(\gamma_1 + \phi_S^{-1}\right)^2}, \tag{37c}$$

where

$$\gamma_0 = \mathrm{Tr}\left[\boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} \right] \tag{38a}$$

$$\gamma_1 = \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Psi}^{-1} \mathbf{g}_{\mathrm{d}} \tag{38b}$$

$$\gamma_2 = \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Psi}^{-1} \mathbf{g}_{\mathrm{d}} \tag{38c}$$

$$\gamma_3 = \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Psi}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Psi}^{-1} \mathbf{g}_{\mathrm{d}}. \tag{38d}$$

Finally, substituting the results from (38) in (32) yields the final expression for the CRB:

$$\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S) = \frac{1}{M} \frac{1}{\gamma_0 - 2\frac{\gamma_3}{\gamma_1} + \frac{\gamma_2^2}{\gamma_1^2} + \frac{2\delta}{\gamma_1^2(\phi_S\gamma_1 + 1)}}, \tag{39}$$

where $\delta \equiv \gamma_3\gamma_1 - \gamma_2^2$.

For the noiseless case the definitions in (38) are simplified to:

$$\gamma_0 = N\phi_R^{-2} \tag{40a}$$

$$\gamma_1 = \phi_R^{-1} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Gamma}^{-1} \mathbf{g}_{\mathrm{d}} \tag{40b}$$

$$\gamma_2 = \phi_R^{-2} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Gamma}^{-1} \mathbf{g}_{\mathrm{d}} \tag{40c}$$

$$\gamma_3 = \phi_R^{-3} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Gamma}^{-1} \mathbf{g}_{\mathrm{d}}. \tag{40d}$$

The expression for the CRB becomes simpler and identical to the CRB derived in [14, Eq. 12]:

$$\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R} = \frac{1}{M} \frac{\phi_R^2}{N-1}. \tag{41}$$

B. CRB for the estimator with BM

In the following section, the FIM and the CRB of ϕ_R w.r.t. the p.d.f. of $\bar{\mathbf{u}}$ is derived. Similarly to (33), the Fisher-information of ϕ_R equals

$$\mathrm{I}^{\bar{\mathbf{u}}}_{RR} = M \ \mathrm{Tr}\left[\boldsymbol{\Phi}_{\mathbf{u}}^{-1} \frac{\partial \boldsymbol{\Phi}_{\mathbf{u}}}{\partial \phi_R} \boldsymbol{\Phi}_{\mathbf{u}}^{-1} \frac{\partial \boldsymbol{\Phi}_{\mathbf{u}}}{\partial \phi_R} \right], \tag{42}$$

and the corresponding CRB is given by

$$\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R} = \left(\mathrm{I}^{\bar{\mathbf{u}}}_{RR}\right)^{-1}. \tag{43}$$

Using (21), one can notice that $\boldsymbol{\Phi}_{\mathbf{u}}^{-1} = \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Psi} \mathbf{B}\right)^{-1}$. Substituting it in (42), the FIM now reads:

$$\mathrm{I}^{\bar{\mathbf{u}}}_{RR} = M \ \mathrm{Tr}\left[\left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Psi} \mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma} \mathbf{B} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Psi} \mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma} \mathbf{B} \right]. \tag{44}$$

To derive the expression for the CRB, the following identity presented in (30) may be used (the identity is proved in Appendix A):

$$\mathbf{B} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Psi} \mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{H}} = \boldsymbol{\Psi}^{-1} - \frac{\boldsymbol{\Psi}^{-1} \mathbf{g}_{\mathrm{d}} \mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Psi}^{-1}}{\mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \boldsymbol{\Psi}^{-1} \mathbf{g}_{\mathrm{d}}}. \tag{45}$$

Using the cyclic property of the trace operation, the Fisher-information from (44) may be written as

$$\mathrm{I}^{\bar{\mathbf{u}}}_{RR} = M \mathrm{Tr}\left[\mathbf{B} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Psi} \mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma} \mathbf{B} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Psi} \mathbf{B}\right)^{-1} \mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma} \right]. \tag{46}$$

Substituting the identity from (30) in (46), and using the definitions in (38), the Fisher information can be obtained as

$$\mathrm{I}^{\bar{\mathbf{u}}}_{RR} = M\left(\gamma_0 - 2\frac{\gamma_3}{\gamma_1} + \frac{\gamma_2^2}{\gamma_1^2}\right). \tag{47}$$

Finally, employing the definition of the CRB in (43) yields the expression for the CRB:

$$\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R} = \frac{1}{M}\frac{1}{\gamma_0 - 2\frac{\gamma_3}{\gamma_1} + \frac{\gamma_2^2}{\gamma_1^2}}. \tag{48}$$

Note that the expressions of the CRBs in (39) and (48) are similar besides the additional component in the denominator of (39), namely $\frac{2\delta}{\gamma_1^2(\phi_S\gamma_1+1)}$. For the noiseless case, it can be shown that $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$ is identical to $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}$ in (41).

C. Comparison of the CRBs

In the following section, the CRBs in (39) and (48) are compared, to determine which of which provides the more accurate estimate in the minimum MSE sense.

Although $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$ is independent of ϕ_S and $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$ depends on ϕ_S, the CRBs may be evaluated as a function of ϕ_S.

The following observations can be made w.r.t. $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$ and $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$:

1) Compared to $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$, $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$ has an additional component in the denominator that is equal to $\frac{2\delta}{\gamma_1^2(\phi_S\gamma_1+1)}$.
   Since γ_1 is positive[1] and ϕ_S is positive, the sign of the additional component depends only on δ. In Appendix B, it is proven that $\delta \geq 0$. Accordingly, it can be concluded that $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$ is always smaller than $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$.
2) The additional component $\frac{2\delta}{\gamma_1^2(\phi_S\gamma_1+1)}$ is inversely proportional to ϕ_S. Hence, it can be deduced that $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$ is monotonically increasing as a function of ϕ_S. Moreover, the limit of $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$ as ϕ_S goes to infinity is

$$\lim_{\phi_s\to\infty} \mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S) = \mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}. \tag{49}$$

Finally, an illustration of $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$ and $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$ as a function of ϕ_S is depicted in Fig. 1.

Some relevant conclusions from the behavior of the CRBs can be drawn:

- Since, theoretically, $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S) \leq \mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$ consistently, the ML estimation of ϕ_R without the blocking operation has, asymptotically, a lower MSE than the ML estimation of ϕ_R with the blocking operation.
- When the signal to reverberation plus noise ratio (SRNR) decreases (i.e. ϕ_S becomes smaller), the performance difference increases. This probably occurs since low levels of speech renders the blocking operation unnecessary. However, in high SRNR cases, the blocking operation becomes more effective.
- In case where $\phi_S = 0$ (i.e. the blocking action is entirely unnecessary), the CRB for the joint estimator is simplified to

$$\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S = 0) = \frac{1}{M}\frac{1}{\left(\gamma_0 - 2\frac{\gamma_3}{\gamma_1} + \frac{\gamma_2^2}{\gamma_1^2}\right) + \frac{2\delta}{\gamma_1^2}}. \tag{50}$$

  Thus, even in cases where the blocking operation is unnecessary, if δ equals zero the CRBs are equal.
- In Appendix B, it is shown that δ equals zero when: 1) the ambient noise level equals zero, or 2) the spatial fields of the ambient noise and the reverberation are identical. In such cases both CRBs are identical.
- For large inter-microphone distances and/or high frequencies, the reverberant sound field tends to be spatially white. If the ambient noise is also spatially white, it can be shown that both CRBs are identical.

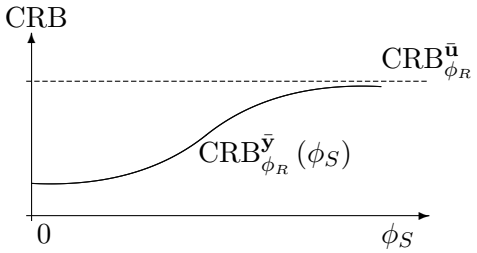

Fig. 1. Illustration of $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}(\phi_S)$ and $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$ as a function of ϕ_S.

V. Performance Evaluation

In this section, an experimental comparison between the reverberation PSD estimators in Sections III-A and III-B is presented. First, a Monte-Carlo experiment was carried out, to investigate and confirm the theoretical results of the paper. Secondly, the reverberation PSD was estimated from real reverberant and noisy signals using the ML estimator and other competing estimators. The average log-error between the estimated reverberation PSDs and the oracle reverberation PSDs is presented. Thirdly, a dereverberation and noise reduction algorithm

[1] $\mathbf{\Psi}$ consists of PSD matrices and therefore is a positive-definite matrix. Thus, according to (38b), γ_1 is always positive.

was applied using estimates of the two estimators (and some other competing estimators). The output signal was compared with the direct-path signals in terms of perceptual evaluation of speech quality (PESQ) score and log-spectral distance (LSD) result.

A. CRB Comparison Using Monte-Carlo Simulations

This section provides several Monte-Carlo simulations to demonstrate the theoretical results obtained in the previous sections.

In each simulation, the influence of only one parameter is examined while the other parameters remain fixed. N microphone signals were synthetically created with M i.i.d. snapshots. Pure-tone signals with frequency $f = 800$ Hz were used throughout this section. A uniform linear array was simulated. The DTF $\mathbf{g}_{\mathrm{d}}$ was assumed to be a pure delay system

$$G_{\mathrm{d},i} = \exp\left(-\iota 2\pi f \tau \cdot i\right), \tag{51}$$

where $\iota = \sqrt{-1}$ and τ is the time difference of arrival (TDOA) between two neighboring microphones. The TDOA depends on the direction of arrival (DOA), $\tau = \frac{d\cos(\theta)}{c}$, where θ is the angle of arrival (from 0° to 180°), d is the inter-microphones distance and c is the sound velocity ($c = 340\frac{\mathrm{m}}{\mathrm{sec}}$). The spatial coherence matrix $\mathbf{\Gamma}$ was set as a diffused noise field, $\mathbf{\Gamma}_{ij} = \mathrm{sinc}\left(2\pi f d|i-j|/c\right)$. The noise PSD matrix was set as $\mathbf{\Phi_v} = \phi_V\,\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix. The values of the nominal parameter set are presented in Table I.

Parameter	Nominal value
Snapshot number M	100
Microphone number N	4
Speech PSD ϕ_S	0.5
Rev. PSD ϕ_R	0.5
Noise PSD ϕ_V	0.5
Inter-microphone distance d	0.06 m
Frequency f	800 Hz
Direction of arrival	90°
Iterations number L	10

TABLE I
NOMINAL PARAMETER-SET FOR THE SIMULATED EXPERIMENTS.

In each simulation, 2000 Monte-Carlo experiments were executed. In each experiment, ϕ_R was estimated according to Algorithm 1 and 2. The MSE of the reverberation PSD was computed.

Four measurements were calculated: 1) the theoretical CRB in (39) refers to the joint ML estimator which circumvents the application of the blocking operation; 2) the MSE of the estimates using Algorithm 1; 3) the theoretical CRB in (48) refers to the ML estimator which use the blocking operation; 4) the MSE of the estimates using Algorithm 2. In the following sections, the four measurements are presented vs. the values of M, ϕ_S, ϕ_R, ϕ_V, N, d, θ and f.

1) CRBs and MSEs as a function of the number of snapshots M: Figure 2 shows the CRBs and the MSEs of the reverberation level estimates as a function of the snapshots number M. The snapshots number varies from 5 to 500 snapshots (it should be noted that since in practice the reverberation is highly time-varying, this high number of snapshot can only be used in this artificial experiment). It can be verified that the theoretical CRBs curves match well with the MSE curves, especially when the number of snapshots is large. In a low number of snapshots, the MSEs curves are even lower than the CRBs curves. This result may be attributed to the lower and upper delimitations described in Algorithm 1 and 2. As expected, the CRBs and MSEs curves decrease as the number of snapshots increase.

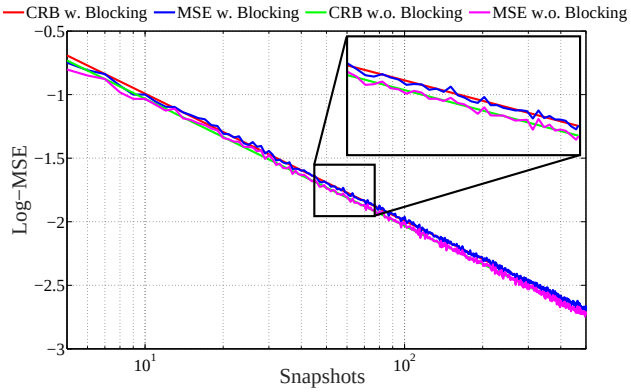

Fig. 2. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of the number of snapshots M.

2) CRBs and MSEs as a function of ϕ_S: In the second experiment, the oracle ϕ_S was set to various values, such that the SRNR varies from -20 dB to 15 dB where $\mathrm{SRNR} \equiv 10\log\left(\frac{\phi_S}{\phi_R+\phi_V}\right)$. Figure 3 shows the CRBs and the MSEs of the reverberation level estimates as a function of the SRNR. It can be verified that the theoretical CRBs curves match with the MSE curves and follow the theoretical contours depicted in Fig. 1. In low SRNR there is a difference between the CRBs while in high SRNR the CRBs are identical.

3) CRBs and MSEs as a function of ϕ_R: In the third experiment, the oracle ϕ_R was set to various values, such that the signal-to-reverberant ratio (SRR) varies from -20 dB to $+20$ dB, where $\mathrm{SRR} \equiv 10\log\left(\frac{\phi_S}{\phi_R}\right)$. The signal-to-noise ratio (SNR) is 0 dB, where $\mathrm{SNR} \equiv 10\log\left(\frac{\phi_S}{\phi_V}\right)$. Figure 4 shows the CRBs and the MSEs

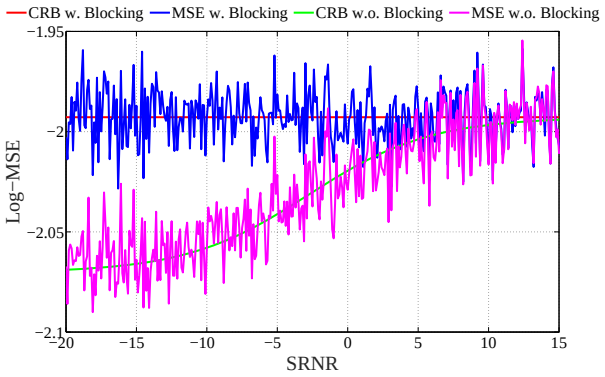

Fig. 3. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of the SRNR.

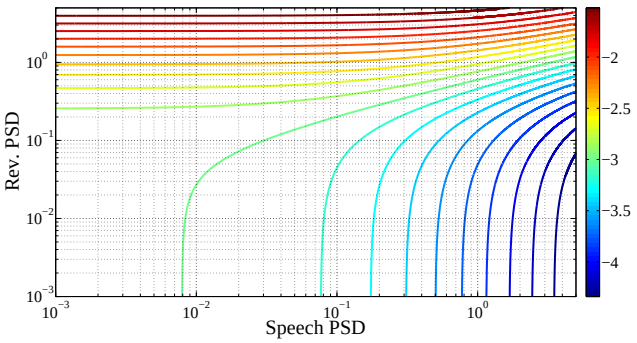

Fig. 5. The difference between the CRBs, D (52), as a function of ϕ_S and ϕ_R.

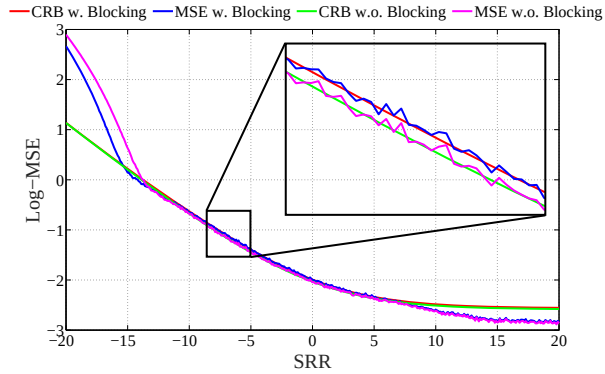

Fig. 4. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of the SRR.

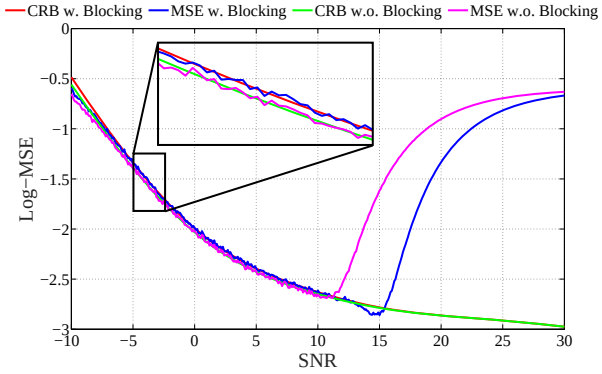

Fig. 6. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of the SNR with $L = 10$.

of the reverberation level estimates as a function of the SRR. In the middle range, the MSE curves match the CRB curves. In high ϕ_R values (i.e., low SRR), the MSE of both estimators are higher than the CRB curves. In low ϕ_R values (i.e., high SRR), the MSEs are smaller than the CRBs. These effects often occur when the oracle value of desired parameter is very small or very large relative to the other parameters. We have no theoretical explanation for these effects.

In Fig. 5, the difference between $\mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R}$ and $\mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}$,

$$\mathrm{D} \equiv \mathrm{CRB}^{\bar{\mathbf{u}}}_{\phi_R} - \mathrm{CRB}^{\bar{\mathbf{y}}}_{\phi_R}, \tag{52}$$

is presented as a function of ϕ_S and ϕ_R. It can be verified that as long ϕ_S decreases or ϕ_R increases the difference becomes higher.

4) CRBs and MSEs as a function of ϕ_V: In the fourth experiment, the noise power ϕ_V was set to various values, such that the SNR varies from -10 dB to $+30$ dB. The SRR is 0 dB. Figure 6 shows the CRBs and the MSEs of the reverberation PSD estimates as a function of the SNR.

In high ϕ_V values (i.e., low SNR), both the MSEs are lower than the CRB curves. In the middle range, the MSE curves match the CRB curves. In low ϕ_V values (i.e., high SNR), the MSEs are much higher than the CRB curves. Moreover, the MSE for Algorithm 2 outperforms the MSE of Algorithm 1. The CRB curves in high SNR case are identical as shown in (41). Apparently, in high SNR cases, the parameters update rule converges slowly [24], especially when two parameters are estimated. Therefore, it is worthwhile to increase the iteration number L in such cases. Figure 7 shows the CRBs and the MSEs of the reverberation PSD estimates as a function of the SNR where $L = 20$ iterations of the Newton search were applied (Nominally, $L = 10$ iterations were applied). It can be verified that the MSEs are now well matched with the CRBs.

5) CRBs and MSEs as a function of the microphone number N: In the fifth experiment, the microphone number is varied from 2 microphones to 20. Figure 8 shows the CRBs and the MSEs of the reverberation PSD estimates as a function of N.

The MSEs and CRBs curves are decreased as the

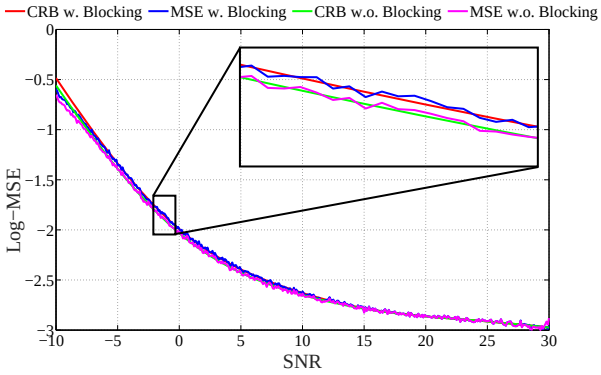

Fig. 7. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of the SNR with $L = 20$.

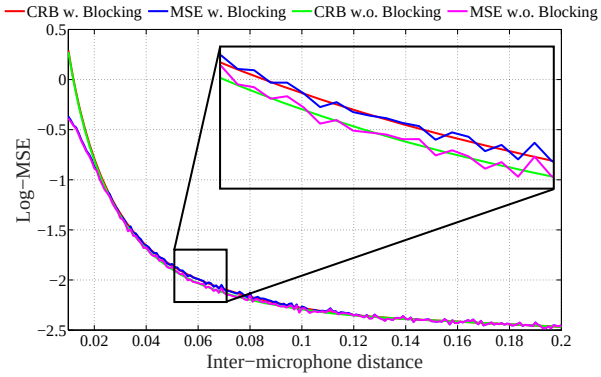

Fig. 9. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of d.

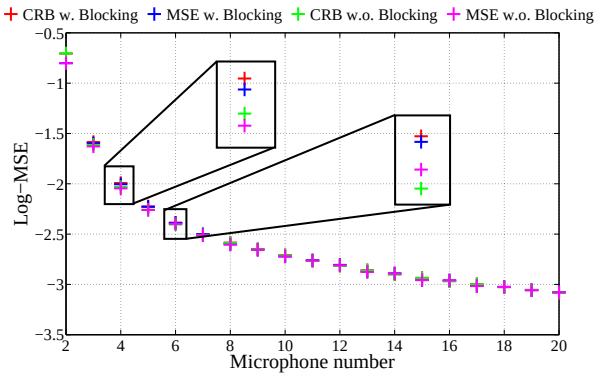

Fig. 8. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of N.

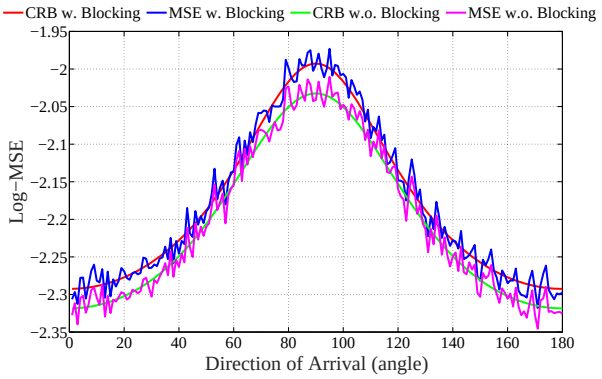

Fig. 10. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of θ.

microphone number is increased. Conceptually, as long as the amount of data is increased the MSE is decreased. Moreover, since the blocking stage blocks only one dimension, the difference between the CRB curves is decreased as long as the dimensions number of the data is increased. There is a difference between the estimators when $3-7$ microphones are used.

6) CRBs and MSEs as a function of the inter-microphones distance d: In the sixth experiment, four microphones was taken. The microphone inter distance d was set to the range $0.01-0.2$ m. Figure 9 depicts the CRBs and the MSEs of the reverberation PSD estimates as a function of d. There is a noticeable difference between the estimators only in the range $0.03-0.12$ m. Generally, the MSEs and the CRBs are decreases as d increases. At large inter-microphone distances the diffuse field behaves as a spatially white noise field. In our case, the fields of the reverberation and the ambient noise are therefore identical. As was mentioned in Sec. IV-C, the CRBs are identical in this case.

7) CRBs and MSEs as a function of the DOA: In the seventh experiment, various directions of arrivals were examined. The angle of arrival θ was set in the range $1^\circ-180^\circ$. Figure 10 depicts the CRBs and the MSEs of the reverberation PSD estimates as a function of θ. It can be observed that the estimator without blocking achieves a lower CRB and MSE compared to the estimator with blocking for all the directions of arrival. Interestingly, in endfire scenarios, the CRBs and the MSEs decreases.

8) CRBs and MSEs as a function of the frequency f: In the sixth experiment, the frequency was set in the range $0-4000$ Hz. Figure 11 shows the CRBs and the MSEs of the reverberation PSD estimates as a function of f. There is a difference between the estimators in the frequency range $400-1300$ Hz. The MSEs decreases as the frequency increases. At high frequencies the diffuse field resembles a spatially white noise field. As the reverberation and ambient noise fields are identical, also the CRBs are identical.

B. Performance Evaluation with Measured Room Impulse Responses

The performance of the two MLEs is now examined using real reverberant and noisy signals. As in [13],

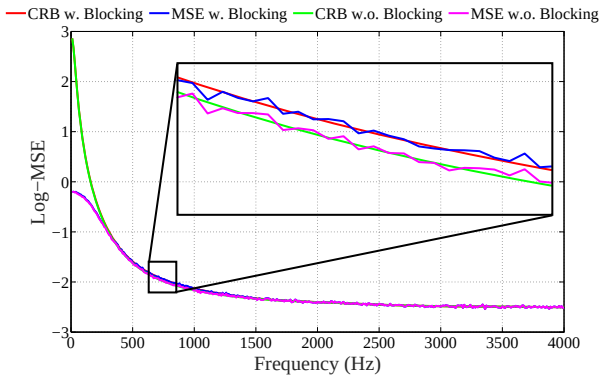

Fig. 11. The CRBs and the MSEs of the reverberation level estimates of the two MLE schemes as a function of f.

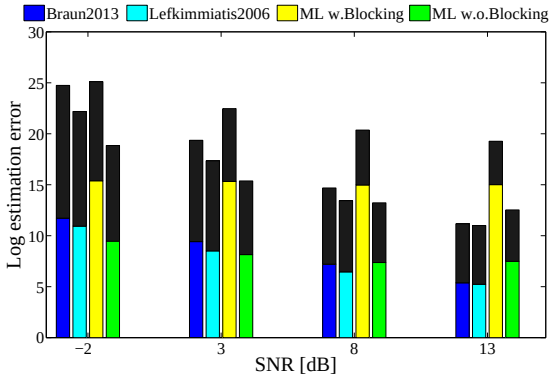

Fig. 12. Log-errors of the two ML reverberation PSD level estimators in comparison with Braun2013 [8] and Lefkimmiatis2006 [26]. The upper part of each bar represents the underestimation error, while the lower part represents the overestimation error.

the performance of the two estimators is evaluated by: 1) examining the log-error between the estimated value of ϕ_R^{ML} and the true reverberation PSD, obtained by convolving the anechoic speech signal with the late component of the acoustic impulse response; and 2) utilizing the estimated PSD levels ϕ_R^{ML} in a speech dereverberation and noise reduction task.

1) Experimental setup: The experiments consist of reverberant signals plus directional noise with various SNR levels. Spatially white (sensor) noise was also added, with power 20 dB lower than the directional noise power. Anechoic speech signals were convolved by room impulse responses (RIRs), downloaded from an open-source database recorded in our lab. Details about the database and RIRs identification method can be found in [25]. Reverberation time was set by adjusting the room panels, and was measured to be approximately $T_{60} = 0.61$ s. The SRR was measured to be approximately -5 dB. The spatial PSD matrix $\mathbf{\Phi_v}$ was estimated using periods in which the desired speech source was inactive. An oracle voice activity detector was used for this task. The loudspeaker was positioned in front of a four microphone linear array, such that the steering vector was set to $\mathbf{g}_{\mathrm{d}} = \left[\begin{array}{cccc} 1 & 1 & 1 & 1 \end{array}\right]^{\mathrm{T}}$. The inter-distances between the microphones were $[3,\ 8,\ 3]$ cm. The sampling frequency was 16 kHz, the frame length of the short-time Fourier transform (STFT) was 32 ms with 8 ms between successive time-frames (i.e., 75% overlap). We have also set $\epsilon = 10^{-10}$. The number of iterations was set to $L = 10$. In cases where the a posteriori SNR $\frac{\frac{1}{N}\mathbf{y}^{\mathrm{H}}\mathbf{y}}{\phi_V}$ was at least 60 dB, the high-SNR estimator (17) was used. All measures were computed by averaging the results obtained using 50 sentences, 4–8 s long, evenly distributed between female and male speakers.

Practically, due to the non-stationarity of the signals, we have substituted the sliding window averaging (9) by recursive averaging that was used also in our previous work [13], i.e.,

$$\mathbf{R}_{\mathbf{y}}(m) = \alpha\, \mathbf{R}_{\mathbf{y}}(m-1) + (1-\alpha)\, \mathbf{y}(m)\mathbf{y}^{\mathrm{H}}(m), \quad (53)$$

where $0 \leq \alpha < 1$ is a smoothing factor. The smoothing factor α was set to 0.7, i.e. only few frames are involved in the smoothing. In each time-index m, $\mathbf{R}_{\mathbf{y}}(m)$ in (53) was used for estimating $\phi_S(m)$ and $\phi_R(m)$.

2) Accuracy of the ML estimator: The performance of the two ML estimators was compared to other two existing estimators in terms of log-errors between the estimated PSD and the oracle PSD: 1) the estimator in [8], denoted henceforth as Braun2013; 2) the estimator in [26][2], denoted henceforth as Lefkimmiatis2006.

In Fig. 12, the results for several SNR levels are depicted.

The mean log-errors between the estimated PSD levels and the oracle PSD levels are presented. In order to calculate the oracle PSD levels, the anechoic speech was filtered with the reverberation tails of the RIRs. The reverberation tails were set to start 2 ms after the arrival time of the direct-path. To reduce the variance of the oracle PSD, the mean value of the oracle PSDs over all microphones was computed. The result bars are split to distinguish between underestimation errors and overestimation errors. It is evident that the joint estimator outperforms the other three competing estimators in terms of overall log-errors for -2 dB, 3 dB, 8 dB and 13 dB SNR values. For the higher SNR of

[2]Note that the algorithm in [26] aims to estimate the noise variance given the received signals PSD matrix while the noise coherence matrix is assumed to be known. In our implementation, we treat the reverberation as an additive noise, and we estimate its level while we subtract the ambient noise PSD matrix from the received signals PSD matrix.

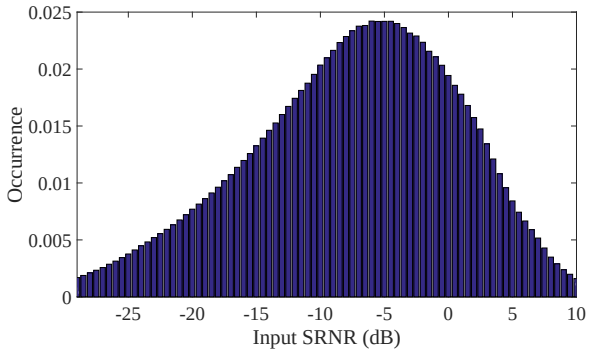

Fig. 13. Occurrence of the various SRNRs in the time-frequency bins

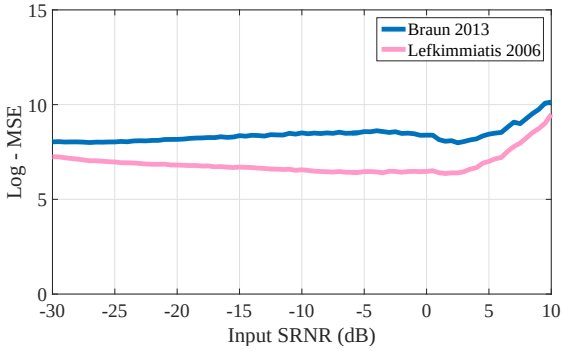

Fig. 15. Mean log-errors of Braun2013 and Lefkimmiatis2006 vs. the SRNR axis.

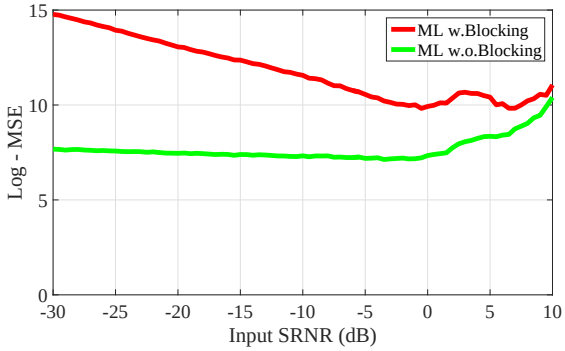

Fig. 14. Mean log-errors of the two ML late reverberation PSD estimators vs. the SRNR axis.

13 dB, Lefkimmiatis2006 and Braun2013 outperform the proposed ML estimators.

We also examined the mean log-errors between the estimated PSD levels and the oracle PSD levels for various values of SRNR. We computed the true SRNR for each time-frequency bin and calculated the mean log-errors for each SRNR separately. In Fig. 13 the occurrence of the various SRNRs in the time-frequency bins is presented. It can be seen that the most common SRNRs are found in the range -10 dB to 0 dB. In Fig. 14, the mean log-errors results are presented vs. the SRNR level for the blocking based ML estimator and the non-blocking based ML estimator. It can be seen that for low SRNR the joint ML estimator outperforms the blocking-ML estimator, but for high SRNR the blocking-ML estimator becomes better. It can be deduced that for most highly occurring SRNRs the joint estimator outperforms blocking-ML estimator. This result is inline with the trends presented in the theoretical derivation in Section IV-C. In Fig. 15, the mean log-errors results are presented vs. the SRNR level for Braun2013 and Lefkimmiatis2006.

3) Dereverberation performance: The performance of the proposed estimator was also examined by utilizing the estimated PSDs in a joint dereverberation and noise reduction task. The estimated PSDs were used to compute the multi-channel Wiener filter presented in [4]. The multichannel Wiener filter (MCWF) was designed to jointly suppress the power of the total interference (e.g. the reverberation and the noise) by deriving the MMSE estimator of the direct path. The MCWF was implemented by a two-stage approach: a minimum variance distortionless response beamformer followed by a corresponding post filter. The performance of the dereverberation algorithm was evaluated in terms of two objective measures, commonly used in the speech enhancement community, namely PESQ [27] and LSD. The clean reference for evaluation in all cases was the anechoic speech signal filtered by only the direct path of the RIR.

In Tables II and III, the performance measures for several input SNR levels are depicted. The joint estimator outperforms all competing estimators with respect to the LSD measures while the estimator that uses the blocking outperforms all competing estimators with respect to the PESQ measures.The PESQ seems to mainly reflect reverberation reduction rather than speech distortion. The estimator that uses the blocking slightly overestimates the reverberation level and hence results in more reverberation reduction. The dereverberation and noise reduction results are given for a specific case just to present a usage of the reverberation power. The differences between the performance of the two variants are indeed small. Note that the oracle PSD of the reverberation level does not yield a significantly better result either.

VI. Conclusions

In this work, two ML estimators of the reverberation PSD level in a noisy environment were compared. The first one jointly estimates the reverberation and the anechoic speech PSDs. The second one first blocks the

SNR	-2 dB	3 dB	8 dB	13 dB
Unprocessed	1.33	1.53	1.74	1.86
Oracle ϕ_R	1.91	2.04	2.11	2.15
Braun2013 [8]	1.78	1.96	2.07	2.13
Lefkimmiatis2006 [26]	1.81	1.98	2.08	2.13
ML with blocking	**1.91**	**2.06**	**2.14**	**2.17**
ML without blocking	1.83	1.97	2.06	2.12

TABLE II
PESQ SCORES FOR THE MCWF [4] USING VARIOUS ESTIMATORS.

SNR	-2 dB	3 dB	8 dB	13 dB
Unprocessed	9.42	8.08	6.91	6.14
Oracle ϕ_R	5.93	5.47	5.17	4.95
Braun2013 [8]	6.27	5.65	5.30	5.10
Lefkimmiatis2006 [26]	6.10	5.58	5.27	5.08
ML with blocking	6.13	5.55	5.20	4.98
ML without blocking	**6.09**	**5.51**	**5.19**	**4.96**

TABLE III
LSD SCORES FOR THE MCWF [4] USING VARIOUS ESTIMATORS.

speech component and then estimates the reverberation level PSD using the outputs of the blocking matrix. The CRB expressions associated with the two ML estimators were derived and compared. The CRB associated with the joint estimator is proven to be lower than the CRB associated with the estimator that uses the blocking stage. An experimental study computed the MSE curves of the two estimators vs. the various problem parameters. Monte-Carlo simulations shows that the MSE is congruous with the CRB curves. The ML estimators were applied to real recordings in order to estimate the reverberation PSD levels. The estimated levels were utilized to implement a recently proposed dereverberation and noise reduction algorithm. The obtained results are presented and compared with the results of obtained by other competing algorithms from the literature.

APPENDIX A

In this appendix, the following identity is proven:

$$\mathbf{B}\left(\mathbf{B}^{\mathrm{H}}\mathbf{\Psi}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{H}} = \mathbf{\Psi}^{-1} - \frac{\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}\mathbf{g}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Psi}^{-1}}{\mathbf{g}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}}. \quad \text{(A.54)}$$

Note that the only condition for the identity is $\mathbf{B}^{\mathrm{H}}\mathbf{g}_{\mathrm{d}} = \mathbf{0}$. In order to prove the above identity, the well-known criterion of the MVDR-beamformer (BF) is used in the following. Regarding the signal model in (3), the MVDR-BF that minimizes the power of the interference (i.e. the reverberation and the noise) while maintaining distortionless response of the direct path is defined by [21]

$$\mathbf{h}_{\mathrm{MVDR}} = \underset{\mathbf{h}}{\operatorname{argmin}}\, \mathbf{h}^{\mathrm{H}}\mathbf{\Psi}\mathbf{h} \;\text{ s.t. }\; \mathbf{h}^{\mathrm{H}}\mathbf{g}_{\mathrm{d}} = F, \quad \text{(A.55)}$$

where F is the required response. The solution for this optimization problem is

$$\mathbf{h}_{\mathrm{MVDR}} = \frac{\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}}{\mathbf{g}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}}F^{*}. \quad \text{(A.56)}$$

The MVDR-BF can be implemented using a generalized sidelobe canceller (GSC) structure [21], [28]:

$$\mathbf{h}_{\mathrm{MVDR}} = \mathbf{h}_0 - \mathbf{B}\mathbf{h}_{\mathrm{NC}}, \quad \text{(A.57)}$$

where $\mathbf{h}_0$ is the fixed beamformer (FBF) satisfying $\mathbf{h}_0^{\mathrm{H}}\mathbf{g}_{\mathrm{d}} = F$ and $\mathbf{h}_{\mathrm{NC}}$ is the noise canceller (NC) that is responsible for mitigating the residual interference at the output of the FBF,

$$\mathbf{h}_{\mathrm{NC}} = \left(\mathbf{B}^{\mathrm{H}}\mathbf{\Psi}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{\Psi}\mathbf{h}_0. \quad \text{(A.58)}$$

A common choice for $\mathbf{h}_0$ is $\mathbf{h}_0 = \frac{\mathbf{g}_{\mathrm{d}}}{||\mathbf{g}_{\mathrm{d}}||^2}$ [21]. Note that for such a choice, the branches of the GSC (i.e. $\mathbf{h}_0$ and $-\mathbf{B}\mathbf{h}_{\mathrm{NC}}$) are *orthogonal*.

In cases where the desired response F is set as $F = G_{d,i}$ (where $G_{d,i}$ is the i-th component of $\mathbf{g}_{\mathrm{d}}$), the FBF can be set as $\mathbf{h}_0 = \mathbf{I}_i$ (where $\mathbf{I}_i$ is the i-th column of the identity matrix $\mathbf{I}$), since $\mathbf{I}_i^{\mathrm{H}}\mathbf{g}_{\mathrm{d}} = G_{d,i}$. Note that by this selection, the MVDR decomposes into two vectors which are not necessarily *orthogonal*. In previous work [4], we elaborate about the benefits that can be gained from such a structure.

Denote the MVDR-BF that satisfies $F = G_{d,i}$ as $\mathbf{h}_{\mathrm{MVDR},i}$, the standard MVDR expression (A.56) and its corresponding GSC implementation (A.57) (which uses $\mathbf{h}_0 = \mathbf{I}_i$) are proved to be identical [21], i.e.,

$$\begin{aligned}\mathbf{h}_{\mathrm{MVDR},i} &= \frac{\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}}{\mathbf{g}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}}G_{d,i}^{*} \\ &= \mathbf{I}_i - \mathbf{B}\left(\mathbf{B}^{\mathrm{H}}\mathbf{\Psi}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{\Psi}\mathbf{I}_i. \quad \text{(A.59)}\end{aligned}$$

Concatenating all the MVDR and the GSC expressions for each $F = G_{d,i}$, the following is obtained:

$$\begin{aligned}\left[\begin{array}{ccc}\mathbf{h}_{\mathrm{MVDR},1} & \ldots & \mathbf{h}_{\mathrm{MVDR},N}\end{array}\right] &= \frac{\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}}{\mathbf{g}_{\mathrm{d}}^{\mathrm{H}}\mathbf{\Psi}^{-1}\mathbf{g}_{\mathrm{d}}}\mathbf{g}_{\mathrm{d}}^{\mathrm{H}} \\ &= \mathbf{I} - \mathbf{B}\left(\mathbf{B}^{\mathrm{H}}\mathbf{\Psi}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{H}}\mathbf{\Psi}\mathbf{I}. \quad \text{(A.60)}\end{aligned}$$

Multiplying from the right both sides of the equation above by $\mathbf{\Psi}^{-1}$ concludes the proof of the identity (A.54).

APPENDIX B

In this appendix, it is proven that $\delta \equiv \gamma_3\gamma_1 - \gamma_2^2$ is a positive number. In the following, a detailed expression for δ is derived. Since γ_1, γ_2 and γ_3 consist of $\mathbf{\Psi}^{-1}$ (see (38b)-(38d)), some transformations may be made to $\mathbf{\Psi}^{-1}$ as elaborated in the sequel.

First, the noise PSD matrix is whitened by decomposing $\mathbf{\Phi_v} = \phi_V \mathbf{C}\mathbf{C}^{\mathrm{H}}$, where ϕ_V is the noise level, $\mathbf{C}\mathbf{C}^{\mathrm{H}}$ is the normalized[3] PSD matrix of the noise and $\mathbf{C}$ is an invertible matrix[4]. Therefore, from (36), $\mathbf{\Psi}^{-1}$ can now be expressed as

$$\mathbf{\Psi}^{-1} = \mathbf{C}^{-H}\left(\phi_R \mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}^{-H} + \phi_V \mathbf{I}\right)^{-1}\mathbf{C}^{-1}. \quad \text{(B.61)}$$

Now, the EVD is applied to $\mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}^{-\mathrm{H}}$, by defining

$$\mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}^{-\mathrm{H}} = \mathbf{V}^{\mathrm{H}}\mathbf{\Lambda}\mathbf{V}. \quad \text{(B.62)}$$

$\mathbf{\Psi}^{-1}$ now reads

$$\mathbf{\Psi}^{-1} = \mathbf{C}^{-\mathrm{H}}\mathbf{V}^{\mathrm{H}}\mathbf{\Upsilon}^{-1}\mathbf{V}\mathbf{C}^{-1}, \quad \text{(B.63)}$$

where $\mathbf{\Upsilon} \equiv \phi_R\mathbf{\Lambda} + \phi_V\mathbf{I}$.

Define $\mathbf{d} \equiv \mathbf{V}\mathbf{C}^{-1}\mathbf{g}_{\mathrm{d}}$ and using (B.63), the definitions in (38b)-(38d) may be re-expressed as

$$\gamma_1 = \mathbf{d}^{\mathrm{H}}\mathbf{\Upsilon}^{-1}\mathbf{d} \quad \text{(B.64a)}$$
$$\gamma_2 = \mathbf{d}^{\mathrm{H}}\mathbf{\Upsilon}^{-1}\mathbf{\Lambda}\mathbf{\Upsilon}^{-1}\mathbf{d} \quad \text{(B.64b)}$$
$$\gamma_3 = \mathbf{d}^{\mathrm{H}}\mathbf{\Upsilon}^{-1}\mathbf{\Lambda}\mathbf{\Upsilon}^{-1}\mathbf{\Lambda}\mathbf{\Upsilon}^{-1}\mathbf{d}. \quad \text{(B.64c)}$$

Since $\mathbf{\Lambda}$ and $\mathbf{\Upsilon}^{-1}$ are diagonal matrixes, γ_1, γ_2 and γ_3 can be again re-expressed

$$\gamma_1 = \sum_i \frac{|\mathbf{d}_i|^2}{\phi_R\lambda_i + \phi_V}, \quad \text{(B.65a)}$$
$$\gamma_2 = \sum_i \frac{|\mathbf{d}_i|^2 \cdot \lambda_i}{\left(\phi_R\lambda_i + \phi_V\right)^2}, \quad \text{(B.65b)}$$
$$\gamma_3 = \sum_i \frac{|\mathbf{d}_i|^2 \cdot \lambda_i^2}{\left(\phi_R\lambda_i + \phi_V\right)^3}, \quad \text{(B.65c)}$$

where $\mathbf{d}_i$ are the elements of $\mathbf{d}$ and λ_i are the diagonal elements of $\mathbf{\Lambda}$.

Substituting the results from (B.65) in δ yields:

$$\delta = \sum_i \frac{|\mathbf{d}_i|^2 \cdot \lambda_i^2}{\left(\phi_R\lambda_i + \phi_V\right)^3} \sum_j \frac{|\mathbf{d}_j|^2}{\phi_R\lambda_j + \phi_V} - \sum_i \frac{|\mathbf{d}_i|^2 \cdot \lambda_i}{\left(\phi_R\lambda_i + \phi_V\right)^2} \sum_j \frac{|\mathbf{d}_j|^2 \cdot \lambda_j}{\left(\phi_R\lambda_j + \phi_V\right)^2}. \quad \text{(B.66)}$$

[3]such that its trace equals N.

[4]since $\mathbf{\Phi_v}$ is symmetric and usually full-rank, $\mathbf{C}$ can be found using the eigenvalue decomposition (EVD) of $\mathbf{\Phi_v}$.

Computing common denominator for the above components and executing several algebraic steps yields

$$\delta = \sum_i \sum_j \frac{|\mathbf{d}_i|^2|\mathbf{d}_j|^2\left(\lambda_i^2 - \lambda_i\lambda_j\right)}{\left(\phi_R\lambda_i + \phi_V\right)^3\left(\phi_R\lambda_j + \phi_V\right)^2}. \quad \text{(B.67)}$$

For double-summation, a useful identity may be used, which is obtained by exchanging the indexes of the summations:

$$\sum_i \sum_j f(i,j) = \frac{1}{2}\sum_i \sum_j f(i,j) + \frac{1}{2}\sum_j \sum_i f(j,i), \quad \text{(B.68)}$$

where $f(j,i)$ is some function of i and j. Using the above identity in (B.67) yields

$$\delta = \frac{1}{2}\sum_i \sum_j \frac{|\mathbf{d}_i|^2|\mathbf{d}_j|^2\left(\lambda_i^2 - \lambda_i\lambda_j\right)}{\left(\phi_R\lambda_i + \phi_V\right)^3\left(\phi_R\lambda_j + \phi_V\right)^2} + \frac{1}{2}\sum_j \sum_i \frac{|\mathbf{d}_j|^2|\mathbf{d}_i|^2\left(\lambda_j^2 - \lambda_j\lambda_i\right)}{\left(\phi_R\lambda_j + \phi_V\right)^3\left(\phi_R\lambda_i + \phi_V\right)^2}. \quad \text{(B.69)}$$

Finally, after computing common denominator and a few algebraic steps, δ is proved to be a positive number since all its components are positive[5],

$$\delta = \frac{\phi_V}{2}\sum_i \sum_j \frac{|\mathbf{d}_i|^2|\mathbf{d}_j|^2\left(\lambda_i - \lambda_j\right)^2}{\left(\phi_R\lambda_j + \phi_V\right)^3\left(\phi_R\lambda_i + \phi_V\right)^3} \geq 0. \quad \text{(B.70)}$$

Some conclusions may be drawn from the final expression of δ: 1) if ϕ_V equals zero (i.e., noiseless case), δ equals zero; or 2) if λ_i are identical for all i, δ equals zero(which occurs when $\mathbf{C}^{-1}\mathbf{\Gamma}\mathbf{C}^{-\mathrm{H}} = \mathbf{I}$, namely the spatial fields of the ambient noise and the reverberation are identical).In cases when δ equals zero, both CRB expressions (with and without blocking) become identical.

REFERENCES

[1] A. Kjellberg, "Effects of reverberation time on the cognitive load in speech communication : Theoretical considerations," *Noise and Health*, vol. 7, no. 25, pp. 11–21, 2004.

[2] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer, 2010.

[3] E. A. P. Habets, "Single- and multi-microphone speech dereverberation using spectral enhancement," Ph.D. Thesis, Technische Universiteit Eindhoven, Jun. 2007.

[4] 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*, pp. 240–251, Feb. 2015.

[5] 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. 770–773, Sep. 2009.

[5]Note that since λ_i are eigenvalues of a positive definite matrix, they are necessarily positive.

[6] 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, 1984.

[7] U. Kjems and J. Jensen, "Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 295–299.

[8] S. Braun and E. A. P. Habets, "Dereverberation in noisy environments using reference signals and a maximum likelihood estimator," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013, pp. 1–5.

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

[10] 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 (WASPAA), New-Paltz, NY, USA, Oct.*, 2015.

[11] A. Kuklasinski, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*, 2014, pp. 61–65.

[12] H. Ye and R. D. DeGroat, "Maximum likelihood DOA estimation and asymptotic Cramér-Rao bounds for additive unknown colored noise," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 938–949, 1995.

[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 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shngehai, China, Mar.*, 2016.

[14] J. Jensen and M. S. Pedersen, "Analysis of beamformer directed single-channel noise reduction system for hearing aid applications," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, Apr.*, 2015.

[15] A. Kuklasinski, S. Doclo, T. Gerkmann, S. Holdt Jensen, and J. Jensen, "Multi-channel PSD estimators for speech dereverberation-a theoretical and experimental comparison," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 91–95.

[16] E. A. P. Habets and J. Benesty, "A two-stage beamforming approach for noise reduction and dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 945–958, 2013.

[17] E. A. P. Habets, "Towards multi-microphone speech dereverberation using spectral enhancement and statistical reverberation models," in *Proceedings Asilomar Conference on Signals, Systems and Computers*, 2008, pp. 806–810.

[18] N. Dal Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," *Signal Processing*, vol. 15, no. 1, pp. 43–56, 1988.

[19] E. A. P. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," *The Journal of the Acoustical Society of America*, vol. 122, pp. 3464–3470, Dec. 2007.

[20] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[21] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Acoustics, Speech and Signal Processing,*, vol. 49, no. 8, pp. 1614–1626, 2001.

[22] W. J. Bangs, *Array processing with generalized beamformers*. Ph.D. dissertation, Yale University, 1972.

[23] H. Hung and M. Kaveh, "On the statistical sufficiency of the coherently averaged covariance matrix for the estimation of the parameters of wideband sources," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 12, 1987, pp. 33–36.

[24] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 550–563, 2010.

[25] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 313–317.

[26] S. Lefkimmiatis, D. Dimitriadis, and P. Maragos, "An optimum microphone array post-filter for speech applications." in *Proceedings of Interspeech*, 2006.

[27] 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," Feb. 2001.

[28] L. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas Propagagation*, vol. 30, no. 1, pp. 27–34, 1982.

Ofer Schwartz received his B.Sc. (Cum Laude) and M.Sc. degrees in Electrical Engineering from Bar-Ilan University, Israel in 2010 and 2013, respectively. He is now pursuing his Ph.D. in Electrical Engineering at the Speech and Signal Processing laboratory of the Faculty of Engineering at Bar-Ilan University. His research interests include statistical signal processing and in particular dereverberation and noise reduction using microphone arrays and speaker localization and tracking.

Sharon Gannot (S'92-M'01-SM'06) received his B.Sc. degree (summa cum laude) from the Technion Israel Institute of Technology, Haifa, Israel in 1986 and the M.Sc. (cum laude) and Ph.D. degrees from Tel-Aviv University, Israel in 1995 and 2000 respectively, all in Electrical Engineering. In 2001 he held a post-doctoral position at the department of Electrical Engineering (ESAT-SISTA) at K.U.Leuven, Belgium. From 2002 to 2003 he held a research and teaching position at the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. Currently, he is a Full Professor at the Faculty of Engineering, Bar-Ilan University, Israel, where he is heading the Speech and Signal Processing laboratory and the Signal Processing Track.

Prof. Gannot is the recipient of Bar-Ilan University outstanding lecturer award for 2010 and 2014. He is also a co-recipient of seven best paper awards.

Prof. Gannot has served as an Associate Editor of the EURASIP Journal of Advances in Signal Processing in 2003-2012, and as an Editor of several special issues on Multi-microphone Speech Processing of the same journal. He has also served as a guest editor of ELSEVIER Speech Communication and Signal Processing journals. Prof. Gannot has served as an Associate Editor of IEEE Transactions on Speech, Audio and Language Processing in 2009-2013. Currently, he is a Senior Area Chair of the same journal. He also serves as a reviewer of many IEEE journals and conferences.

Prof. Gannot is a member of the Audio and Acoustic Signal Processing (AASP) technical committee of the IEEE since Jan., 2010. Since Jan. 2017, he serves as the committee chair. He is also a member of the Technical and Steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) since 2005 and was the general co-chair of IWAENC held at Tel-Aviv, Israel in August 2010. Prof. Gannot has served as the general co-chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in October 2013. Prof. Gannot was selected (with colleagues) to present a tutorial sessions in ICASSP 2012, EUSIPCO 2012, ICASSP 2013 and EUSIPCO 2013. Prof. Gannot research interests include multi-microphone speech processing and specifically distributed algorithms for ad hoc microphone arrays for noise reduction and speaker separation; dereverberation; single microphone speech enhancement and speaker localization and tracking.

Emanuël A.P. Habets (S'02-M'07-SM'11) is an Associate Professor at the International Audio Laboratories Erlangen (a joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg and Fraunhofer IIS), and Head of the Spatial Audio Research Group at Fraunhofer IIS, Germany. He received the B.Sc. degree in electrical engineering from the Hogeschool Limburg, The Netherlands, in 1999, and the M.Sc. and Ph.D. degrees in electrical engineering from the Technische Universiteit Eindhoven, The Netherlands, in 2002 and 2007, respectively.

From 2007 until 2009, he was a Postdoctoral Fellow at the Technion - Israel Institute of Technology and at the Bar-Ilan University, Israel. From 2009 until 2010, he was a Research Fellow in the Communication and Signal Processing Group at Imperial College London, U.K.

His research activities center around audio and acoustic signal processing, and include spatial audio signal processing, spatial sound recording and reproduction, speech enhancement (dereverberation, noise reduction, echo reduction), and sound localization and tracking.

Dr. Habets was a member of the organization committee of the 2005 International Workshop on Acoustic Echo and Noise Control (IWAENC) in Eindhoven, The Netherlands, a general co-chair of the 2013 International Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in New Paltz, New York, and general co-chair of the 2014 International Conference on Spatial Audio (ICSA) in Erlangen, Germany. He was a member of the IEEE Signal Processing Society Standing Committee on Industry Digital Signal Processing Technology (2013-2015), and a Guest Editor for the IEEE Journal of Selected Topics in Signal Processing and the EURASIP Journal on Advances in Signal Processing. He is the recipient, with S. Gannot and I. Cohen, of the 2014 IEEE Signal Processing Letters Best Paper Award. Currently, he is a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, vice-chair of the EURASIP Special Area Team on Acoustic, Sound and Music Signal Processing, an Associate Editor of the IEEE Signal Processing Letters, and Editor in Chief of the EURASIP Journal on Audio, Speech, and Music Processing.