ML Estimation and CRBs for Reverberation, Speech and Noise PSDs in Rank-Deficient Noise Field

Yaron Laufer, *Student Member, IEEE*, Bracha Laufer-Goldshtein, *Student Member, IEEE*, and
Sharon Gannot, *Senior Member, IEEE*

Abstract—Speech communication systems are prone to performance degradation in reverberant and noisy acoustic environments. Dereverberation and noise reduction algorithms typically require several model parameters, e.g. the speech, reverberation and noise power spectral densities (PSDs). A commonly used assumption is that the noise PSD matrix is known. However, in practical acoustic scenarios, the noise PSD matrix is unknown and should be estimated along with the speech and reverberation PSDs. In this paper, we consider the case of rank-deficient noise PSD matrix, which arises when the noise signal consists of multiple directional noise sources, whose number is less than the number of microphones. We derive two closed-form maximum likelihood estimators (MLEs). The first is a non-blocking-based estimator which jointly estimates the speech, reverberation and noise PSDs, and the second is a blocking-based estimator, which first blocks the speech signal and then jointly estimates the reverberation and noise PSDs. Both estimators are analytically compared and analyzed, and mean square errors (MSEs) expressions are derived. Furthermore, Cramér-Rao Bounds (CRBs) on the estimated PSDs are derived. The proposed estimators are examined using both simulation and real reverberant and noisy signals, demonstrating the advantage of the proposed method compared to competing estimators.

Index Terms—Dereverberation, Noise reduction, Maximum likelihood estimation, Cramér-Rao Bound.

I. Introduction

In many hands-free scenarios, the measured microphone signals are corrupted by additive background noise, which may originate from both environmental sources and from microphone responses. Apart from noise, if the recording takes place in an enclosed space, the recorded signals may also contain multiple sound reflections from walls and other objects in the room, resulting in reverberation. As the level of noise and reverberation increases, the perceived quality and intelligibility of the speech signal deteriorate, which in turn affect the performance of speech communication systems, as well as automatic speech recognition (ASR) systems.

In order to reduce the effects of reverberation and noise, speech enhancement algorithms are required, which aim at recovering the clean speech source from the recorded microphone signals. Speech dereverberation and noise reduction algorithms often require the power spectral densities (PSDs) of the speech, reverberation and noise components. In the multichannel framework, a commonly used assumption (see e.g. in [1]–[10]) is that the late reverberant signal is modelled

Yaron Laufer, Bracha Laufer-Goldshtein and Sharon Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan, 5290002, Israel (e-mail: Yaron.Laufer@biu.ac.il, Bracha.Laufer@biu.ac.il, Sharon.Gannot@biu.ac.il).

as a spatially homogeneous sound field, with a time-varying PSD multiplied by a spherical diffuse time-invariant spatial coherence matrix. As the spatial coherence matrix depends only on the microphone geometry, it can be calculated in advance. However, the reverberation PSD is an unknown parameter that should be estimated. Numerous methods exist for estimating the reverberation PSD. They are broadly divided into two classes, namely non-blocking-based estimators and blocking-based estimators. The non-blocking-based approach jointly estimates the PSDs of the late reverberation and speech. The estimation is carried out using the maximum likelihood (ML) criterion [3], [6] or in the least-squares (LS) sense, by minimizing the Frobenius norm of an error PSD matrix [8]. In the blocking-based method, the desired speech signal is first blocked using a blocking matrix (BM), and then the reverberation PSD is estimated. Estimators in this class are also based on the ML approach [1], [5], [7], [9] or the LS criteria [2], [4].

All previously mentioned methods do not include an estimator for the noise PSD. In [1], [3], a noiseless scenario is assumed. In [2], [4]–[10], the noise PSD matrix is assumed to be known in advance, or that an estimate is available. Typically, the noise PSD matrix is assumed to be time-invariant, and therefore can be estimated during speech-absent periods using a voice activity detector (VAD). However, in practical acoustic scenarios the spectral characteristics of the noise might be time-varying, e.g. when the noise environment includes a background radio or TV, and thus a VAD-based algorithm may fail. Therefore, the noise PSD matrix has to be included in the estimation procedure.

Some papers in the field deal with performance analysis of the proposed estimators. We give a brief review of the commonly used tools to assess the quality of an estimator. Theoretical analysis of estimators typically consists of calculating the bias and the mean square error (MSE), which coincides with the variance for unbiased estimators. The Cramér-Rao Bound (CRB) is an important tool to evaluate the quality of any unbiased estimator, since it gives a lower bound on the MSE. An estimator that is unbiased and attains the CRB, is called *efficient*. The maximum likelihood estimator (MLE) is *asymptotically* efficient [11], namely attains the CRB when the amount of samples is large.

Theoretical analysis of PSD estimators in the noise-free scenario was addressed in [12], [13]. In [12], CRBs were derived for the reverberation and the speech MLEs proposed in [3]. These MLEs are efficient, i.e. attain the CRB for any number of samples. In addition, it was pointed out that the non-

blocking-based reverberation MLE derived in [3] is identical to the blocking-based MLE proposed in [1]. In [13], it was shown that the non-blocking-based reverberation MLE of [3] obtains a lower MSE compared to a noiseless version of the blocking-based LS estimator derived in [2].

In the noisy case, quality assessment was discussed in [9] and [14]. In [9], it was numerically demonstrated that an iterative blocking-based MLE yields lower MSE than the blocking-based LS estimator proposed in [2]. In [14], closed-form CRBs were derived for the two previously proposed MLEs of the reverberation PSD, namely the blocking-based estimator in [5] and the non-blocking-based estimator in [6]. The CRB for the non-blocking-based reverberation estimator was shown to be lower than the CRB for the blocking-based estimator. However, it was shown that in the noiseless case, both reverberation MLEs are identical and both CRBs coincide.

As opposed to previous works, the assumption of known noise PSD matrix is not made in [15]. The noise is modelled as a spatially homogeneous sound field, with a time-varying PSD multiplied by a time-invariant spatial coherence matrix. It is assumed that the spatial coherence matrix of the noise is known in advance, while the time-varying PSD is unknown. Two different estimators were developed, based on the LS method. In the first one, a joint estimator for the speech, noise and late reverberation PSDs was developed. As an alternative, a blocking-based estimator was proposed, in which the speech signal is first blocked by a BM, and then the noise and reverberation PSDs are jointly estimated. This method was further extended in [16] to jointly estimate the various PSDs along with the relative early transfer function (RETF) vector, using an alternating least-squares method. However, the model used in [15] and [16] only fits spatially homogeneous noise fields that are characterized by a full-rank covariance matrix. Moreover, in [9] it was claimed that the ML approach is preferable over the LS estimation procedure.

Recently, a confirmatory factor analysis was used in [17] to jointly estimate the PSDs and the RETF. The noise is modelled as a microphone-self noise, with a time-invariant diagonal PSD matrix. A closed-form solution is not available for this case, thus requiring iterative optimization techniques.

In this paper, we treat the noise PSD matrix as an unknown parameter. We assume that the noise PSD matrix is a *rank-deficient* matrix, as opposed to the spatially homogeneous assumption considered in [15]. This scenario arises when the noise signal consists of a set of directional noise sources, whose number is smaller than the number of microphones. We assume that the positions of the noise sources are fixed, while their spectral PSD matrix is time-varying, e.g. when the acoustic environment includes radio or TV. It should be emphasized that, in contrast to [15] which estimates only a scalar PSD of the noise, in our model the entire spectral PSD matrix of the noise is estimated, and thus the case of multiple non-stationary noise sources, can be handled. We derive closed-form MLEs of the various PSDs, for both the non-blocking-based and the blocking-based methods. The proposed estimators are analytically studied and compared, and the corresponding MSEs expressions are derived. Furthermore,

CRBs for estimating the various PSDs are derived.

An important benefit of considering the rank-deficient noise as a separated problem, is due to the form of the solution. In the ML framework, a closed-form solution exists for the noiseless case [1], [3] but not for the full-rank noise scenario, thus requiring iterative optimization techniques [5], [6], [9] (as opposed to LS method that has closed-form solutions in both cases). However, we show here that when the noise PSD matrix is a rank-deficient matrix, closed-form MLE exists, which yields simpler and faster estimation procedure with low computational complexity, and is not sensitive to local maxima.

The remainder of the paper is organized as follows. Section II presents the problem formulation, and describes the probabilistic model. Section III derives the MLEs for both the non-blocking-based and the blocking-based methods, and Section IV presents the CRB derivation. Section V demonstrates the performance of the proposed estimators by an experimental study based on both simulated data and recorded room impulse responses (RIRs). The paper is concluded in Section VI.

II. PROBLEM FORMULATION

In this section, we formulate the dereverberation and noise reduction problem. Scalars are denoted with regular lowercase letters, vectors are denoted with bold lowercase letters and matrices are denoted with bold uppercase letters. A list of notations used in our derivations is given in Table I.

A. Signal Model

Consider a speech signal received by N microphones, in a noisy and reverberant acoustic environment. We work with the short-time Fourier transform (STFT) representation of the measured signals. Let $k \in [1, K]$ denote the frequency bin index, and $m \in [1, M]$ denote the time frame index. The N-channel observation signal $\mathbf{y}(m, k) \in \mathbb{C}^N$ writes

$$\mathbf{y}(m, k) = \mathbf{x}_e(m, k) + \mathbf{r}(m, k) + \mathbf{u}(m, k), \quad (1)$$

where $\mathbf{x}_e(m, k) \in \mathbb{C}^N$ denotes the direct and early reverberation speech component, $\mathbf{r}(m, k) \in \mathbb{C}^N$ denotes the late reverberation speech component and $\mathbf{u}(m, k) \in \mathbb{C}^N$ denotes the noise. The direct and early reverberation speech component is given by $\mathbf{x}_e(m, k) = \mathbf{g}_d(k)s(m, k)$, where $s(m, k) \in \mathbb{C}$ is the direct and early speech component as received by the first microphone (designated as a reference microphone), and $\mathbf{g}_d(k) = [1, g_{d,2}(k), \cdots, g_{d,N}(k)]^\top \in \mathbb{C}^N$ is the time-invariant relative early transfer function (RETF) vector between the reference microphone and all microphones. In this paper, we follow previous works in the field, e.g. [2], [13]–[15], and neglect the early reflections. Thus, the target signal $s(m, k)$ is approximated as the direct component at the reference microphone, and $\mathbf{g}_d(k)$ reduces to the relative direct-path transfer function (RDTF) vector. It is assumed that the noise signal consists of T noise sources, i.e.

$$\mathbf{u}(m, k) = \mathbf{A}_\mathbf{u}(k)\mathbf{s}_\mathbf{u}(m, k), \quad (2)$$

where $\mathbf{s}_\mathbf{u}(m, k) \in \mathbb{C}^T$ denotes the vector of noise sources and $\mathbf{A}_\mathbf{u}(k) \in \mathbb{C}^{N \times T}$ is the noise acoustic transfer function (ATF)

TABLE I: Nomenclature

Symbol	Meaning
$(\cdot)^{\top}$	transpose
$(\cdot)^{\mathrm{H}}$	conjugate transpose
$(\cdot)^{*}$	complex conjugate
$\lvert\cdot\rvert$	determinant of a matrix
$\mathrm{Tr}[\cdot]$	trace of a matrix
$\otimes$	Kronecker product
$\mathrm{vec}(\cdot)$	stacking the columns of a matrix on top of one another
N	No. of microphones, $n \in \{1, \dots, N\}$
M	No. of time frames, $m \in \{1, \dots, M\}$
K	No. of frequency bins, $k \in \{1, \dots, K\}$
T	No. of noise sources, $t \in \{1, \dots, T\}$
$\mathbf{y}$	Observation signal
s	Direct speech component
$\mathbf{g}_d$	Relative direct-path transfer function
$\mathbf{r}$	Late reverberation component
$\mathbf{u}$	Noise signal
$\mathbf{A_u}$	Noise ATF matrix
$\mathbf{s_u}$	Vector of noise sources
ϕ_S	Speech PSD
ϕ_R	Reverberation PSD
$\mathbf{\Gamma}_R$	Reverberation coherence matrix
$\mathbf{\Psi_u}$	Noise PSD matrix

matrix, assumed to be time-invariant. It is further assumed that $T \leq N - 2$.

B. Probabilistic Model

The speech STFT coefficients are assumed to follow a zero-mean complex Gaussian distribution[1] with a time-varying PSD $\phi_S(m,k)$. Hence, the PDF of the speech writes:

$$p\big(s(m,k); \phi_S(m,k)\big) = \mathcal{N}_c\big(s(m,k); 0, \phi_S(m,k)\big), \quad (3)$$

The late reverberation signal is modelled by a zero-mean complex multivariate Gaussian distribution:

$$p\big(\mathbf{r}(m,k); \mathbf{\Phi_r}(m,k)\big) = \mathcal{N}_c\big(\mathbf{r}(m,k); \mathbf{0}, \mathbf{\Phi_r}(m,k)\big). \quad (4)$$

The reverberation PSD matrix is modelled as a spatially homogeneous and isotropic sound field, with a time-varying PSD, $\mathbf{\Phi_r}(m,k) = \phi_R(m,k)\mathbf{\Gamma}_R(k)$. It is assumed that the time-invariant coherence matrix $\mathbf{\Gamma}_R(k)$ can be modelled by a spherically diffuse sound field [19]:

$$\Gamma_{R,ij}(k) = \mathrm{sinc}\left(\frac{2\pi f_s k}{K}\frac{d_{ij}}{c}\right), \quad (5)$$

where $\mathrm{sinc}(x) = \sin(x)/x$, d_{ij} is the inter-distance between microphones i and j, f_s denotes the sampling frequency and c is the sound velocity.

[1]The multivariate complex Gaussian probability density function (PDF) is given by $\mathcal{N}_c(\mathbf{a}; \boldsymbol{\mu}_a, \mathbf{\Phi}_a) = \frac{1}{\lvert \pi \mathbf{\Phi}_a \rvert} \exp\big(-(\mathbf{a}-\boldsymbol{\mu}_a)^{\mathrm{H}}\mathbf{\Phi}_a^{-1}(\mathbf{a}-\boldsymbol{\mu}_a)\big)$, where $\boldsymbol{\mu}_a$ is the mean vector and $\mathbf{\Phi}_a$ is an Hermitian positive definite complex covariance matrix [18].

The noise sources vector is modelled by a zero-mean complex multivariate Gaussian distribution with a time-varying PSD matrix $\mathbf{\Psi_u}(m,k) \in \mathbb{C}^{T \times T}$:

$$p\big(\mathbf{s_u}(m,k); \mathbf{\Psi_u}(m,k)\big) = \mathcal{N}_c\big(\mathbf{s_u}(m,k); \mathbf{0}, \mathbf{\Psi_u}(m,k)\big). \quad (6)$$

Using (2) and (6), it follows that $\mathbf{u}(m,k)$ has a zero-mean complex multivariate Gaussian distribution with a PSD matrix $\mathbf{\Phi_u}(m,k) \in \mathbb{C}^{N \times N}$, given by

$$\mathbf{\Phi_u}(m,k) = \mathbf{A_u}(k)\mathbf{\Psi_u}(m,k)\mathbf{A_u^{\mathrm{H}}}(k). \quad (7)$$

Note that $\mathrm{rank}(\mathbf{\Phi_u}) = \mathrm{rank}(\mathbf{\Psi_u}) = T \leq N-2$, i.e. $\mathbf{\Phi_u}$ is a rank-deficient matrix. The PDF of $\mathbf{y}(m,k)$ writes

$$p\big(\mathbf{y}(m,k); \mathbf{\Phi_y}(m,k)\big) = \mathcal{N}_c\big(\mathbf{y}(m,k); \mathbf{0}, \mathbf{\Phi_y}(m,k)\big), \quad (8)$$

where $\mathbf{\Phi_y}$ is the PSD matrix of the input signals. Assuming that the components in (1) are independent, $\mathbf{\Phi_y}$ is given by

$$\begin{aligned}\mathbf{\Phi_y}(m,k) &= \phi_S(m,k)\mathbf{g}_d(k)\mathbf{g}_d^{\mathrm{H}}(k) + \phi_R(m,k)\,\mathbf{\Gamma}_R(k) \\ &\quad + \mathbf{A_u}(k)\mathbf{\Psi_u}(m,k)\mathbf{A_u^{\mathrm{H}}}(k). \end{aligned} \quad (9)$$

A commonly used dereverberation and noise reduction technique is to estimate the speech signal using the multichannel minimum mean square error (MMSE) estimator, which yields the multichannel Wiener filter (MCWF), given by [20]:

$$\hat{s}_{\mathrm{MCWF}}(m,k) = \frac{\mathbf{g}_d^{\mathrm{H}}(k)\mathbf{\Phi}_i^{-1}(m,k)}{\mathbf{g}_d^{\mathrm{H}}(k)\mathbf{\Phi}_i^{-1}(m,k)\mathbf{g}_d(k) + \phi_S^{-1}(m,k)}\mathbf{y}(m,k), \quad (10)$$

where

$$\mathbf{\Phi}_i(m,k) \triangleq \phi_R(m,k)\,\mathbf{\Gamma}_R(k) + \mathbf{A_u}(k)\mathbf{\Psi_u}(m,k)\mathbf{A_u^{\mathrm{H}}}(k) \quad (11)$$

denotes the total interference PSD matrix. For implementing (10), we assume that the RDTF vector $\mathbf{g}_d$ and the spatial coherence matrix $\mathbf{\Gamma}_R$ are known in advance. The RDTF depends only on the direction of arrival (DOA) of the speaker and the geometry of the microphone array, and thus it can be constructed based on a DOA estimate. The spatial coherence matrix is calculated using (5), based on the spherical diffuseness assumption. It follows that estimators of the late reverberation, speech and noise PSDs are required for evaluating the MCWF.

It should be noted that estimating directly the complete $N \times N$ time-varying noise PSD matrix $\mathbf{\Phi_u}(m,k)$, along with the speech and reverberation PSDs, is a complex problem, and a closed-form ML solution is not available. In this paper, we rely on the decomposition of the noise PSD matrix in (7), along with the rank-deficiency assumption. By utilizing the time-invariant ATF matrix $\mathbf{A_u}(k)$, we can construct a projection matrix onto the subspace orthogonal to the noise subspace, which generates nulls towards the noise sources. In the following section, we show that this step enables the derivation of closed-form MLEs for the various PSDs.

The noise ATF matrix $\mathbf{A_u}(k)$ is in general not available, since the estimation of the individual ATFs requires that each noise is active separately. Instead, we only assume the availability of a speech-absent segment, denoted by m_0, in which all noise sources are concurrently active. Based on this segment, we compute a basis for the noise subspace

that can be used instead of the unknown ATF matrix. To this end, we compute the noise PSD matrix $\mathbf{\Phi_u}(m_0, k)$, and apply the eigenvalue decomposition (EVD) to the resulting matrix. Recall that the rank of $\mathbf{\Phi_u}$ is T. Accordingly, a T-rank representation of the noise PSD matrix is formed by the computed eigenvalues and eigenvectors:

$$\mathbf{\Phi_u}(m_0, k) = \mathbf{V}(k)\mathbf{\Lambda}(m_0, k)\mathbf{V}^{\mathrm{H}}(k), \qquad (12)$$

where $\mathbf{\Lambda}(m_0, k) \in \mathbb{C}^{T \times T}$ is the eigenvalues matrix (comprised of the non-zero eigenvalues) and $\mathbf{V}(k) = [\mathbf{v}_1(k), \cdots, \mathbf{v}_T(k)] \in \mathbb{C}^{N \times T}$ is the corresponding eigenvectors matrix. $\mathbf{V}(k)$ is a basis that spans the noise ATFs subspace, and thus [21]

$$\mathbf{A_u}(k) = \mathbf{V}(k)\mathbf{G}(k), \qquad (13)$$

where $\mathbf{G}(k) \in \mathbb{C}^{T \times T}$ consists of projection coefficients of the original ATFs on the basis vectors. Substituting (13) into (2) and using (1), yields

$$\mathbf{y}(m, k) = \mathbf{g}_d(k)s(m, k) + \mathbf{r}(m, k) + \mathbf{V}(k)\mathbf{s_v}(m, k), \quad (14)$$

where $\mathbf{s_v}(m, k) \triangleq \mathbf{G}(k)\mathbf{s_u}(m, k)$. It follows that the noise PSD matrix in (7) can be recast as

$$\mathbf{\Phi_u}(m, k) = \mathbf{V}(k)\mathbf{\Psi_v}(m, k)\mathbf{V}^{\mathrm{H}}(k), \qquad (15)$$

where $\mathbf{\Psi_v}(m, k) = \mathbf{G}(k)\mathbf{\Psi_u}(m, k)\mathbf{G}^{\mathrm{H}}(k)$. Using this basis change, the MCWF in (10) is now computed with

$$\mathbf{\Phi}_i(m, k) = \phi_R(m, k)\,\mathbf{\Gamma}_R(k) + \mathbf{V}(k)\mathbf{\Psi_v}(m, k)\mathbf{V}^{\mathrm{H}}(k). \quad (16)$$

As a result, rather than requiring the knowledge of the exact noise ATF matrix, we use $\mathbf{V}$ that is learned from a speech-absent segment. Due to this basis change, we will need to estimate $\mathbf{\Psi_v}$ instead of $\mathbf{\Psi_u}$.

To summarize, estimators of ϕ_R, ϕ_S and $\mathbf{\Psi_v}$ are required. For the sake of brevity, the frame index m and the frequency bin index k are henceforth omitted whenever possible.

III. ML ESTIMATORS

We propose two ML-based methods: (i) Non-blocking-based estimation: Simultaneous ML estimation of the reverberation, speech and noise PSDs; and (ii) Blocking-based estimation: Elimination of the speech PSD using a blocking matrix (BM), and then joint ML estimation of the reverberation and noise PSDs. Both methods are then compared and analyzed.

A. Non-Blocking-Based Estimation

We start with the joint ML estimation of the reverberation, speech and noise PSDs. Based on the short-time stationarity assumption [9], [12], it is assumed that the PSDs are approximately constant across small number of consecutive time frames, denoted by L. We therefore denote $\bar{\mathbf{y}}(m) \in \mathbb{C}^{LN}$ as the concatenation of L previous observations of $\mathbf{y}(m)$:

$$\bar{\mathbf{y}}(m) \triangleq \left[\mathbf{y}^{\top}(m - L + 1), \cdots, \mathbf{y}^{\top}(m)\right]^{\top}. \qquad (17)$$

The set of unknown parameters is denoted by $\boldsymbol{\phi}(m) = \left[\phi_R(m), \phi_S(m), \boldsymbol{\psi}_V^{\top}(m)\right]^{\top}$, where $\boldsymbol{\psi}_V$ is a vector containing all elements of $\mathbf{\Psi_v}$, i.e. $\boldsymbol{\psi}_V = \mathrm{vec}\left(\mathbf{\Psi_v}\right)$. Assuming that the

L consecutive signals in $\bar{\mathbf{y}}$ are i.i.d., the PDF of $\bar{\mathbf{y}}$ writes (see e.g. [14]):

$$p\big(\bar{\mathbf{y}}(m); \boldsymbol{\phi}(m)\big)$$
$$= \left(\frac{1}{\pi^N |\mathbf{\Phi_{\bar y}}(m)|} \exp\left(-\mathrm{Tr}\left[\mathbf{\Phi_{\bar y}^{-1}}(m)\mathbf{R_y}(m)\right]\right)\right)^{L}, \quad (18)$$

where $\mathbf{R_y}$ is the sample covariance matrix, given by

$$\mathbf{R_y}(m) = \frac{1}{L}\sum_{\ell = m - L + 1}^{m} \mathbf{y}(\ell)\mathbf{y}^{\mathrm{H}}(\ell). \qquad (19)$$

The MLE of the set $\boldsymbol{\phi}(m)$ is therefore given by

$$\boldsymbol{\phi}^{\mathrm{ML}, \bar{\mathbf{y}}}(m) = \underset{\boldsymbol{\phi}(m)}{\mathrm{argmax}} \log p\big(\bar{\mathbf{y}}(m); \boldsymbol{\phi}(m)\big). \qquad (20)$$

To the best of our knowledge, for the general noisy scenario this problem is considered as having no closed-form solution. However, we will show that when the noise PSD matrix $\mathbf{\Phi_u}$ is rank-deficient, with $T = \mathrm{rank}(\mathbf{\Phi_u}) \leq N - 2$, a closed-form solution exists. In the following, we present the proposed estimators. The detailed derivations appear in the Appendices.

In Appendix A, it is shown that the MLE of $\phi_R(m)$ is given by:

$$\phi_R^{\mathrm{ML}, \bar{\mathbf{y}}}(m) = \frac{1}{N - (T + 1)}\mathrm{Tr}\big[\mathbf{Q}\mathbf{R_y}(m)\mathbf{\Gamma}_R^{-1}\big], \qquad (21)$$

where $\mathbf{Q} \in \mathbb{C}^{N \times N}$ is given by

$$\mathbf{Q} = \mathbf{I}_N - \mathbf{A}\left(\mathbf{A}^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}, \qquad (22)$$

and $\mathbf{A} \in \mathbb{C}^{N \times (T+1)}$ is the *speech-plus-noise* subspace

$$\mathbf{A} = [\mathbf{g}_d, \mathbf{v}_1, \cdots, \mathbf{v}_T]. \qquad (23)$$

The matrix $\mathbf{Q}$ is a projection matrix onto the subspace orthogonal to the speech-plus-noise subspace. The role of $\mathbf{Q}$ is to block the directions of the desired speech and noise signals, in order to estimate the reverberation level.

Once we obtain the MLE for the late reverberation PSD, the MLEs for the speech and noise PSDs can be computed. In Appendix B, it is shown that the MLE for the speech PSD writes

$$\phi_S^{\mathrm{ML}, \bar{\mathbf{y}}}(m) = \mathbf{w}_s^{\mathrm{H}}\left(\mathbf{R_y}(m) - \phi_R^{\mathrm{ML}, \bar{\mathbf{y}}}(m)\mathbf{\Gamma}_R\right)\mathbf{w}_s, \qquad (24)$$

where $\mathbf{w}_s \in \mathbb{C}^N$ is a minimum variance distortionless response (MVDR) beamformer that extracts the speech signal while eliminating the noise, given by

$$\mathbf{w}_s^{\mathrm{H}} = \frac{\mathbf{g}_d^{\mathrm{H}}\boldsymbol{P}_v^{\perp}\mathbf{\Gamma}_R^{-1}}{\mathbf{g}_d^{\mathrm{H}}\boldsymbol{P}_v^{\perp}\mathbf{\Gamma}_R^{-1}\mathbf{g}_d}, \qquad (25)$$

and $\boldsymbol{P}_v^{\perp} \in \mathbb{C}^{N \times N}$ is a projection matrix onto the subspace orthogonal to the noise subspace, given by

$$\boldsymbol{P}_v^{\perp} = \mathbf{I}_N - \mathbf{\Gamma}_R^{-1}\mathbf{V}\left(\mathbf{V}^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}\mathbf{V}\right)^{-1}\mathbf{V}^{\mathrm{H}}. \qquad (26)$$

Note that

$$\mathbf{w}_s^{\mathrm{H}}\mathbf{g}_d = 1 \;,\; \mathbf{w}_s^{\mathrm{H}}\mathbf{V} = \mathbf{0}. \qquad (27)$$

The estimator in (24) can be interpreted as the variance of the noisy observations minus the estimated variance of the

reverberation, at the output of the MVDR beamformer [12].

In Appendix C, it is shown that the MLE of the noise PSD can be computed with

$$\boldsymbol{\Psi}_{\mathbf{v}}^{\mathrm{ML},\bar{\mathbf{y}}}(m) = \mathbf{W}_{\mathbf{u}}^{\mathrm{H}} \left(\mathbf{R}_{\mathbf{y}}(m) - \phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\boldsymbol{\Gamma}_R \right) \mathbf{W}_{\mathbf{u}}, \quad (28)$$

where $\mathbf{W}_{\mathbf{u}} \in \mathbb{C}^{N \times T}$ is a multi-source linearly constrained minimum variance (LCMV) beamformer that extracts the noise signals while eliminating the speech signal:

$$\mathbf{W}_{\mathbf{u}}^{\mathrm{H}} = \left(\mathbf{V}^{\mathrm{H}} \boldsymbol{P}_g^{\perp} \boldsymbol{\Gamma}_R^{-1} \mathbf{V} \right)^{-1} \mathbf{V}^{\mathrm{H}} \boldsymbol{P}_g^{\perp} \boldsymbol{\Gamma}_R^{-1}, \quad (29)$$

and $\boldsymbol{P}_g^{\perp} \in \mathbb{C}^{N \times N}$ is a projection matrix onto the subspace orthogonal to the speech subspace, given by

$$\boldsymbol{P}_g^{\perp} = \mathbf{I}_N - \frac{\boldsymbol{\Gamma}_R^{-1} \mathbf{g}_d \mathbf{g}_d^{\mathrm{H}}}{\mathbf{g}_d^{\mathrm{H}} \boldsymbol{\Gamma}_R^{-1} \mathbf{g}_d}. \quad (30)$$

Note that

$$\mathbf{W}_{\mathbf{u}}^{\mathrm{H}} \mathbf{g}_d = \mathbf{0} \;,\; \mathbf{W}_{\mathbf{u}}^{\mathrm{H}} \mathbf{V} = \mathbf{I}_T. \quad (31)$$

Interestingly, the projection matrix $\mathbf{Q}$ can be recast as a linear combination of the above beamformers (see Appendix D):

$$\mathbf{Q} = \mathbf{I}_N - \mathbf{g}_d \mathbf{w}_s^{\mathrm{H}} - \mathbf{V}\mathbf{W}_{\mathbf{u}}^{\mathrm{H}}. \quad (32)$$

Using (27), (31) and (32), it can also be noted that $\mathbf{Q}$ is orthogonal to both beamformers

$$\mathbf{w}_s^{\mathrm{H}} \mathbf{Q} = \mathbf{0} \;,\; \mathbf{W}_{\mathbf{u}}^{\mathrm{H}} \mathbf{Q} = \mathbf{0}. \quad (33)$$

In the noiseless case, i.e. when $T = 0$, $\mathbf{Q}$ reduces to

$$\mathbf{Q} = \mathbf{I}_N - \mathbf{g}_d \mathbf{w}_{s_0}^{\mathrm{H}} = \left(\boldsymbol{P}_g^{\perp} \right)^{\mathrm{H}}, \quad (34)$$

where $\mathbf{w}_{s_0}^{\mathrm{H}} = \frac{\mathbf{g}_d^{\mathrm{H}} \boldsymbol{\Gamma}_R^{-1}}{\mathbf{g}_d^{\mathrm{H}} \boldsymbol{\Gamma}_R^{-1} \mathbf{g}_d}$, leading to the same closed-form estimators as in [3, Eq. (7)]:

$$\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m) = \frac{1}{N-1} \mathrm{Tr} \left[\left(\mathbf{I}_N - \mathbf{g}_d \mathbf{w}_{s_0}^{\mathrm{H}} \right) \mathbf{R}_{\mathbf{y}}(m) \boldsymbol{\Gamma}_R^{-1} \right], \quad (35)$$

$$\phi_S^{\mathrm{ML},\bar{\mathbf{y}}}(m) = \mathbf{w}_{s_0}^{\mathrm{H}} \left(\mathbf{R}_{\mathbf{y}}(m) - \phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\boldsymbol{\Gamma}_R \right) \mathbf{w}_{s_0}. \quad (36)$$

B. Blocking-Based Estimation

As a second approach, we first block the speech component using a BM, and then jointly estimate the PSDs of the reverberation and noise. Let $\mathbf{B} \in \mathbb{C}^{N \times (N-1)}$ denote the BM, which satisfies $\mathbf{B}^{\mathrm{H}} \mathbf{g}_d = \mathbf{0}$. The output of the BM is given by

$$\mathbf{z}(m) \triangleq \mathbf{B}^{\mathrm{H}} \mathbf{y}(m) = \mathbf{B}^{\mathrm{H}} \left(\mathbf{r}(m) + \mathbf{u}(m) \right). \quad (37)$$

The PDF of $\mathbf{z}(m) \in \mathbb{C}^{N-1}$ therefore writes:

$$p\left(\mathbf{z}(m); \boldsymbol{\Phi}_{\mathbf{z}}(m) \right) = \mathcal{N}_c \left(\mathbf{z}(m); \mathbf{0}, \boldsymbol{\Phi}_{\mathbf{z}}(m) \right), \quad (38)$$

with the PSD matrix $\boldsymbol{\Phi}_{\mathbf{z}}(m) \in \mathbb{C}^{(N-1) \times (N-1)}$ given by

$$\boldsymbol{\Phi}_{\mathbf{z}}(m) = \mathbf{B}^{\mathrm{H}} \boldsymbol{\Phi}_i(m) \mathbf{B}, \quad (39)$$

where $\boldsymbol{\Phi}_i$ is the total interference matrix, defined in (11). Multiplying (15) from left by $\mathbf{B}^{\mathrm{H}}$ and from right by $\mathbf{B}$, the noise PSD matrix at the output of the BM writes

$$\mathbf{B}^{\mathrm{H}} \boldsymbol{\Phi}_{\mathbf{u}}(m) \mathbf{B} = \widetilde{\mathbf{V}} \boldsymbol{\Psi}_{\mathbf{v}}(m) \widetilde{\mathbf{V}}^{\mathrm{H}}, \quad (40)$$

where $\widetilde{\mathbf{V}} \in \mathbb{C}^{(N-1) \times T}$ is the *reduced* noise subspace:

$$\widetilde{\mathbf{V}} \triangleq \mathbf{B}^{\mathrm{H}} \mathbf{V} = [\tilde{\mathbf{v}}_1, \cdots, \tilde{\mathbf{v}}_T]. \quad (41)$$

Substituting (16) into (39) and using (41), yields

$$\boldsymbol{\Phi}_{\mathbf{z}}(m) = \phi_R(m) \mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma}_R \mathbf{B} + \widetilde{\mathbf{V}} \boldsymbol{\Psi}_{\mathbf{v}}(m) \widetilde{\mathbf{V}}^{\mathrm{H}}. \quad (42)$$

Under this model, the parameter set of interest is $\tilde{\boldsymbol{\phi}}(m) = \left[\phi_R(m), \boldsymbol{\psi}_V^{\top}(m) \right]^{\top}$. Let $\bar{\mathbf{z}} \in \mathbb{C}^{L(N-1)}$ be defined similarly to $\bar{\mathbf{y}}$ in (17), as a concatenation of L i.i.d. consecutive frames. Similarly to $\boldsymbol{\Phi}_{\mathbf{y}}(m)$, it is assumed that $\boldsymbol{\Phi}_{\mathbf{z}}(m)$ is fixed during the entire segment. The PDF of $\bar{\mathbf{z}}$ therefore writes

$$p\left(\bar{\mathbf{z}}(m); \tilde{\boldsymbol{\phi}}(m) \right)$$
$$= \left(\frac{1}{\pi^{N-1}|\boldsymbol{\Phi}_{\mathbf{z}}|} \exp \left(- \mathrm{Tr} \left[\boldsymbol{\Phi}_{\mathbf{z}}^{-1}(m) \mathbf{R}_{\mathbf{z}}(m) \right] \right) \right)^L, \quad (43)$$

where $\mathbf{R}_{\mathbf{z}}(m)$ is given by

$$\mathbf{R}_{\mathbf{z}}(m) = \frac{1}{L} \sum_{\ell=m-L+1}^{m} \mathbf{z}(\ell) \mathbf{z}^{\mathrm{H}}(\ell) = \mathbf{B}^{\mathrm{H}} \mathbf{R}_{\mathbf{y}}(m) \mathbf{B}. \quad (44)$$

The MLE of $\tilde{\boldsymbol{\phi}}(m)$ is obtained by solving:

$$\tilde{\boldsymbol{\phi}}^{\mathrm{ML},\bar{\mathbf{z}}}(m) = \underset{\tilde{\boldsymbol{\phi}}(m)}{\mathrm{argmax}} \log p\left(\bar{\mathbf{z}}(m); \tilde{\boldsymbol{\phi}}(m) \right). \quad (45)$$

To the best of our knowledge, this problem is also considered as having no closed-form solution. Again, we argue that if the noise PSD matrix satisfies $T = \mathrm{rank}(\boldsymbol{\Phi}_{\mathbf{u}}) \leq N - 2$, then we can obtain a closed-form solution. In Appendix E, the following MLE is obtained:

$$\phi_R^{\mathrm{ML},\bar{\mathbf{z}}}(m) = \frac{1}{N-1-T} \mathrm{Tr} \left[\widetilde{\mathbf{Q}} \mathbf{R}_{\mathbf{z}}(m) \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma}_R \mathbf{B} \right)^{-1} \right], \quad (46)$$

where $\widetilde{\mathbf{Q}} \in \mathbb{C}^{(N-1) \times (N-1)}$ is given by

$$\widetilde{\mathbf{Q}} = \mathbf{I}_{N-1} - \widetilde{\mathbf{V}} \left(\widetilde{\mathbf{V}}^{\mathrm{H}} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma}_R \mathbf{B} \right)^{-1} \widetilde{\mathbf{V}} \right)^{-1} \widetilde{\mathbf{V}}^{\mathrm{H}} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma}_R \mathbf{B} \right)^{-1}. \quad (47)$$

After the BM was applied, the remaining role of $\widetilde{\mathbf{Q}}$ is to block the noise signals, in order to estimate the reverberation level. Note that $\mathrm{Tr}\left[\mathbf{Q} \right] = \mathrm{Tr}\left[\widetilde{\mathbf{Q}} \right] = N - 1 - T$.

Given $\phi_R^{\mathrm{ML},\bar{\mathbf{z}}}$, it is shown in Appendix F that the MLE for the noise PSD writes

$$\boldsymbol{\Psi}_{\mathbf{v}}^{\mathrm{ML},\bar{\mathbf{z}}}(m) = \widetilde{\mathbf{W}}_{\mathbf{u}}^{\mathrm{H}} \left(\mathbf{R}_{\mathbf{z}}(m) - \phi_R^{\mathrm{ML},\bar{\mathbf{z}}}(m) \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma}_R \mathbf{B} \right) \right) \widetilde{\mathbf{W}}_{\mathbf{u}}, \quad (48)$$

where $\widetilde{\mathbf{W}}_{\mathbf{u}} \in \mathbb{C}^{(N-1) \times T}$ is a multi-source LCMV beamformer, directed towards the noise signals after the BM, given by

$$\widetilde{\mathbf{W}}_{\mathbf{u}}^{\mathrm{H}} = \left(\widetilde{\mathbf{V}}^{\mathrm{H}} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma}_R \mathbf{B} \right)^{-1} \widetilde{\mathbf{V}} \right)^{-1} \widetilde{\mathbf{V}}^{\mathrm{H}} \left(\mathbf{B}^{\mathrm{H}} \boldsymbol{\Gamma}_R \mathbf{B} \right)^{-1}. \quad (49)$$

Note that with this notation, $\widetilde{\mathbf{Q}}$ in (47) can be recast as

$$\widetilde{\mathbf{Q}} = \mathbf{I}_{N-1} - \widetilde{\mathbf{V}} \widetilde{\mathbf{W}}_{\mathbf{u}}^{\mathrm{H}}. \quad (50)$$

Since $\widetilde{\mathbf{W}}_{\mathbf{u}}^{\mathrm{H}} \widetilde{\mathbf{V}} = \mathbf{I}_T$, it also follows that

$$\widetilde{\mathbf{W}}_{\mathbf{u}}^{\mathrm{H}} \widetilde{\mathbf{Q}} = \mathbf{0}. \quad (51)$$

Also, in Appendix G it is shown that

$$\mathbf{W_u} = \mathbf{B}\widetilde{\mathbf{W}_u}, \tag{52}$$

namely the LCMV of (29), used in the non-blocking-based approach, can be factorized into two stages: The first is a BM that blocks the speech signal, followed by a modified LCMV, which recovers the noise signals at the output of the BM.

C. Comparing the MLEs

In this section, the obtained blocking-based and non-blocking-based MLEs are compared. We will use the following identity, that is proved in [14, Appendix A]:

$$\mathbf{B}\left(\mathbf{B}^{\mathrm{H}}\mathbf{\Gamma}_R\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{H}} = \mathbf{\Gamma}_R^{-1} - \frac{\mathbf{\Gamma}_R^{-1}\mathbf{g}_d\mathbf{g}_d^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}}{\mathbf{g}_d^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}\mathbf{g}_d}. \tag{53}$$

Substituting (30) into (53) yields

$$\mathbf{B}\left(\mathbf{B}^{\mathrm{H}}\mathbf{\Gamma}_R\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{H}} = \boldsymbol{P}_g^{\perp}\mathbf{\Gamma}_R^{-1}. \tag{54}$$

1) Comparing the reverberation PSD estimators: First, we compare the reverberation PSD estimators in (21) and (46). Substituting (50) into (46) and then using (41), (44), (52) and (53), yields the following equation:

$$\phi_R^{\mathrm{ML},\bar{\mathbf{z}}} = \frac{1}{N-1-T}$$
$$\times \mathrm{Tr}\left[\left(\mathbf{I}_N - \frac{\mathbf{g}_d\mathbf{g}_d^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}}{\mathbf{g}_d^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}\mathbf{g}_d}\right)\left(\mathbf{I}_N - \mathbf{V}\mathbf{W}_u^{\mathrm{H}}\right)\mathbf{R_y}(m)\mathbf{\Gamma}_R^{-1}\right]. \tag{55}$$

Using (32) and noting that $\mathbf{g}_d^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}\mathbf{Q} = \mathbf{0}$, yields (21). It follows that both estimators are identical:

$$\phi_R^{\mathrm{ML},\bar{\mathbf{y}}} = \phi_R^{\mathrm{ML},\bar{\mathbf{z}}}. \tag{56}$$

It should be noted that in [12], [14] the two MLEs of the reverberation PSD were shown to be identical in the noiseless case. Here we extend this result to the noisy case, when the noise PSD matrix is a rank-deficient matrix.

2) Comparing the noise PSD estimators: The noise PSD estimators in (28) and (48) are now compared. Substituting (44) into (48) and then using (52) and (56), yields the same expression as in (28), and therefore

$$\mathbf{\Psi_v}^{\mathrm{ML},\bar{\mathbf{y}}} = \mathbf{\Psi_v}^{\mathrm{ML},\bar{\mathbf{z}}}. \tag{57}$$

D. MSE Calculation

In the sequel, the theoretical performance of the proposed PSD estimators is analyzed. Since the non-blocking-based and the blocking-based MLEs were proved in section III-C to be identical for both reverberation and noise PSDs, it suffices to analyze the non-blocking-based MLEs.

1) Theoretical performance of the reverberation PSD estimators: It is well known that for an unbiased estimator, the MSE is identical to the variance. We therefore start by showing that the non-blocking-based MLE in (21) is unbiased. Using (19), the expectation of (21) writes

$$\mathbb{E}\left(\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) = \frac{1}{N-1-T}\mathrm{Tr}\left[\mathbf{\Phi_y}(m)\mathbf{\Gamma}_R^{-1}\mathbf{Q}\right]. \tag{58}$$

Then, we use the following property (see (84d)):

$$\mathbf{\Gamma}_R^{-1}\mathbf{Q} = \phi_R(m)\mathbf{\Phi_y}^{-1}(m)\mathbf{Q}, \tag{59}$$

to obtain

$$\mathbb{E}\left(\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) = \frac{\phi_R(m)}{N-1-T}\mathrm{Tr}\left[\mathbf{Q}\right] = \phi_R(m). \tag{60}$$

It follows that the reverberation MLE is unbiased, and thus the MSE is identical to the variance. Using the i.i.d. assumption, the variance of the non-blocking-based MLE in (21) is given by

$$\mathrm{var}\left(\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) = \frac{1}{L}\frac{1}{(N-1-T)^2}\mathrm{var}(\mathbf{y}^{\mathrm{H}}(m)\mathbf{\Gamma}_R^{-1}\mathbf{Q}\mathbf{y}(m)). \tag{61}$$

In order to simplify (61), the following identity is used. For a positive definite Hermitian form $\mathbf{x}^{\mathrm{H}}\mathbf{Z}\mathbf{x}$, where $\mathbf{x} \sim \mathcal{N}_c(\mathbf{0}, \mathbf{\Phi_x})$ and $\mathbf{Z}$ a Hermitian matrix, the variance is given by [11, p. 513, Eq. (15.30)]:

$$\mathrm{var}\left(\mathbf{x}^{\mathrm{H}}\mathbf{Z}\mathbf{x}\right) = \mathrm{Tr}\left[\mathbf{\Phi_x}\mathbf{Z}\mathbf{\Phi_x}\mathbf{Z}\right]. \tag{62}$$

Since $\mathbf{y}(m) \sim \mathcal{N}_c(\mathbf{0}, \mathbf{\Phi_y}(m))$ and $\mathbf{\Gamma}_R^{-1}\mathbf{Q}$ is a Hermitian matrix (note that $\mathbf{Q}^{\mathrm{H}} = \mathbf{\Gamma}_R^{-1}\mathbf{Q}\mathbf{\Gamma}_R$), we obtain

$$\mathrm{var}\left(\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) = \frac{1}{L}\frac{1}{(N-1-T)^2}$$
$$\times \mathrm{Tr}\left[\mathbf{\Phi_y}(m)\mathbf{\Gamma}_R^{-1}\mathbf{Q}\mathbf{\Phi_y}(m)\mathbf{\Gamma}_R^{-1}\mathbf{Q}\right]. \tag{63}$$

Finally, using (59) and (84b), the variance writes

$$\mathrm{var}\left(\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) = \frac{1}{L}\frac{\phi_R^2(m)}{N-1-T}. \tag{64}$$

Note that in the noiseless case, namely $T = 0$, the variance reduces to the one derived in [12], [13].

2) Theoretical performance of the noise PSD estimators: Using (19), (9) and (31), and based on the unbiasedness of $\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)$, it can be shown that $\mathbf{\Psi_v}^{\mathrm{ML},\bar{\mathbf{y}}}(m)$ in (28) is an unbiased estimator of $\mathbf{\Psi_v}(m)$.

Next, we calculate the variance of the diagonal terms of $\mathbf{\Psi_v}^{\mathrm{ML},\bar{\mathbf{y}}}(m)$. To this end, we write the (i, j) entry of $\mathbf{\Psi_v}^{\mathrm{ML},\bar{\mathbf{y}}}(m)$ in (28) as

$$\left[\mathbf{\Psi_v}^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right]_{ij} = \mathbf{w}_i^{\mathrm{H}}\left(\mathbf{R_y}(m) - \phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\mathbf{\Gamma}_R\right)\mathbf{w}_j, \tag{65}$$

for $i, j = 1, \ldots, T$, where $\mathbf{w}_i$ is the ith column of the matrix $\mathbf{W_u}$ in (29). Using a partitioned matrix to simplify $\mathbf{w}_i$, it can be shown that

$$\mathbf{w}_i^{\mathrm{H}} = \frac{\mathbf{v}_i^{\mathrm{H}}\boldsymbol{P}_{\bar{\mathbf{V}}_i}^{\perp}\boldsymbol{P}_g^{\perp}\mathbf{\Gamma}_R^{-1}}{\mathbf{v}_i^{\mathrm{H}}\boldsymbol{P}_{\bar{\mathbf{V}}_i}^{\perp}\boldsymbol{P}_g^{\perp}\mathbf{\Gamma}_R^{-1}\mathbf{v}_i}, \tag{66}$$

where $\bar{\mathbf{V}}_i \in \mathbb{C}^{N\times T-1}$ is composed of all the vectors in $\mathbf{V}$ except $\mathbf{v}_i$, i.e. $\bar{\mathbf{V}}_i = [\mathbf{v}_1, \cdots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \cdots, \mathbf{v}_T]$, and $\boldsymbol{P}_{\bar{\mathbf{V}}_i}^{\perp} \in \mathbb{C}^{N\times N}$ is the corresponding projection matrix onto the subspace orthogonal to $\bar{\mathbf{V}}_i$

$$\boldsymbol{P}_{\bar{\mathbf{V}}_i}^{\perp} = \mathbf{I}_N - \boldsymbol{P}_g^{\perp}\mathbf{\Gamma}_R^{-1}\bar{\mathbf{V}}_i\left(\bar{\mathbf{V}}_i^{\mathrm{H}}\boldsymbol{P}_g^{\perp}\mathbf{\Gamma}_R^{-1}\bar{\mathbf{V}}_i\right)^{-1}\bar{\mathbf{V}}_i^{\mathrm{H}}. \tag{67}$$

It can be verified that $\mathbf{w}_i^{\mathrm{H}}\mathbf{v}_j = \delta_{ij}$. Denote the diagonal terms

of $\boldsymbol{\Psi_v}$ by $\{\psi_i\}_{i=1}^T$. In Appendix H, it is shown that

$$
\begin{aligned}
&\mathrm{var}\left(\psi_i^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) \\
&= \frac{\psi_i^2(m)}{L}\left[\left(\frac{1+\xi_i(m)}{\xi_i(m)}\right)^2 + \frac{1}{N-1-T}\frac{1}{\xi_i^2(m)}\right], \quad (68)
\end{aligned}
$$

where $\xi_i(m)$ is defined as the *noise-to-reverberation ratio* at the output of $\mathbf{w}_i$:

$$
\xi_i(m) \triangleq \frac{\psi_i(m)}{\mathbf{w}_i^{\mathrm{H}}\phi_R(m)\boldsymbol{\Gamma}_R\mathbf{w}_i} = \frac{\psi_i(m)}{\phi_R(m)}\mathbf{v}_i^{\mathrm{H}}\boldsymbol{P}_{\bar{\mathbf{V}}_i}^{\perp}\boldsymbol{P}_g^{\perp}\boldsymbol{\Gamma}_R^{-1}\mathbf{v}_i. \tag{69}
$$

Note that

$$
\lim_{\xi_i(m)\to\infty}\mathrm{var}\left(\psi_i^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) = \frac{\psi_i^2(m)}{L}. \tag{70}
$$

3) Theoretical performance of the speech PSD estimator: Using (19), (9) and (27) and based on the unbiasedness of $\phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)$, it can be shown that $\phi_S^{\mathrm{ML},\bar{\mathbf{y}}}(m)$ is an unbiased estimator of $\phi_S(m)$. In a similar manner to (68), the variance of (24) can be shown to be

$$
\begin{aligned}
&\mathrm{var}\left(\phi_S^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) \\
&= \frac{\phi_S^2(m)}{L}\left[\left(\frac{1+\varepsilon(m)}{\varepsilon(m)}\right)^2 + \frac{1}{N-1-T}\frac{1}{\varepsilon^2(m)}\right], \quad (71)
\end{aligned}
$$

where $\varepsilon(m)$ is defined as the *signal-to-reverberation ratio* at the output of $\mathbf{w}_s$:

$$
\varepsilon(m) \triangleq \frac{\phi_S(m)}{\mathbf{w}_s^{\mathrm{H}}\phi_R(m)\boldsymbol{\Gamma}_R\mathbf{w}_s} = \frac{\phi_S(m)}{\phi_R(m)}\mathbf{g}_d^{\mathrm{H}}\boldsymbol{P}_v^{\perp}\boldsymbol{\Gamma}_R^{-1}\mathbf{g}_d. \tag{72}
$$

Note that

$$
\lim_{\varepsilon(m)\to\infty}\mathrm{var}\left(\phi_S^{\mathrm{ML},\bar{\mathbf{y}}}(m)\right) = \frac{\phi_S^2(m)}{L}. \tag{73}
$$

In the noiseless case, i.e. $T=0$, (71) becomes identical to the variance derived in [12], [13].

IV. CRB Derivation

In this section, we derive the CRB on the variance of any unbiased estimator of the various PSDs.

A. CRB for the Late Reverberation PSD

In Appendix I, it is shown that the CRB on the reverberation PSD writes

$$
\mathrm{CRB}\left(\phi_R\right) = \frac{1}{L}\frac{\phi_R^2}{N-(T+1)}. \tag{74}
$$

The resulting CRB is identical to the MSE derived in (64), and thus the proposed MLE is an efficient estimator.

B. CRB for the Speech and Noise PSDs

The CRB on the speech PSD is identical to the MSE derived in (71), as outlined in Appendix J. The CRB on the noise PSD can be derived similarly. We conclude that the proposed PSD estimators are efficient.

V. Experimental Study

In this section, the proposed MLEs are evaluated in a synthetic Monte-Carlo simulation as well as on measurements of a real room environment. In Section V-A, a Monte-Carlo simulation is conducted in which signals are generated synthetically based on the assumed statistical model. The sensitivity of the proposed MLEs is examined with respect to the various model parameters, and the MSEs of the proposed MLEs are compared to the corresponding CRBs. In Section V-B, the proposed estimators are examined in a real room environment, by utilizing them for the task of speech dereverberation and noise reduction using the MCWF.

A. Monte-Carlo Simulation

1) Simulation Setup: In order to evaluate the accuracy of the proposed estimators, synthetic data was generated according to the signal model in (1), by simulating L i.i.d. snapshots of single-tone signals, having a frequency of $f = 2000$ Hz. The signals are captured by a uniform linear array (ULA) with N microphones, and inter-distance d between adjacent microphones. The desired signal component s was drawn according to a complex Gaussian distribution with a zero-mean and a PSD ϕ_S. The RDTF is given by

$$
\mathbf{g}_d = \exp\left(-j2\pi f\boldsymbol{\tau}\right), \tag{75}
$$

where $\boldsymbol{\tau} \in \mathbb{R}^N$ is the time difference of arrival (TDOA) w.r.t. the reference microphone, given by $\boldsymbol{\tau} = \frac{d\sin(\theta)}{c} \times [0, \cdots, N-1]^{\top}$, and θ is the DOA, defined as the broadside angle measured w.r.t. the perpendicular to the array. The reverberation component $\mathbf{r}$ was drawn according to a complex Gaussian distribution with a zero-mean and a PSD matrix $\boldsymbol{\Phi_r} = \phi_R\boldsymbol{\Gamma}_R$, where $\boldsymbol{\Gamma}_R$ is modelled as an ideal spherical diffuse sound field, given by $\Gamma_{R,ij} = \mathrm{sinc}\left(2\pi f\frac{d|i-j|}{c}\right)$, $i, j \in 1, \ldots, N$. The noise component was constructed as $\mathbf{u} = \mathbf{A_u}\mathbf{s_u}$ where $\mathbf{s_u}$ denotes the T noise sources, drawn according to a zero-mean complex Gaussian distribution with a random PSD matrix $\boldsymbol{\Psi_u}$, and $\mathbf{A_u}$ is an $N \times T$ random ATF matrix. For the estimation procedure, $\mathbf{V}$ is extracted by applying the EVD to a set of noisy training samples, generated with different $\boldsymbol{\Psi_u}$.

In the sequel, we examine the proposed estimators and bounds as a function of the model parameters. Specifically, the influence of the following parameters is examined: i) number of snapshots L; ii) reverberation PSD value ϕ_R; iii) speech PSD value ϕ_S; and iv) noise power P_u, which is defined as the Frobenius norm of the noise PSD matrix, i.e. $P_u \triangleq \|\boldsymbol{\Phi_u}\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}\left[\boldsymbol{\Phi_u^{\mathrm{H}}\Phi_u}\right]}$. In each experiment, we changed the value of one parameter, while keeping the rest fixed. The nominal values of the parameters are presented in Table II.

For each scenario, we carried out 1000 Monte-Carlo trials. The reverberation PSD ϕ_R was estimated in each trial with both (21) and (46), the noise PSD was estimated with both (28) and (48) and the speech PSD was estimated with (24). The accuracy of the estimators was evaluated using the normalized mean square error (nMSE), by averaging over the Monte-Carlo trials and normalizing w.r.t. the square of the corresponding PSD value. For each quantity, the corresponding normalized

(a) Increasing L (from left to right)

(b) Decreasing ϕ_R (from left to right)

(c) Increasing ϕ_S (from left to right)

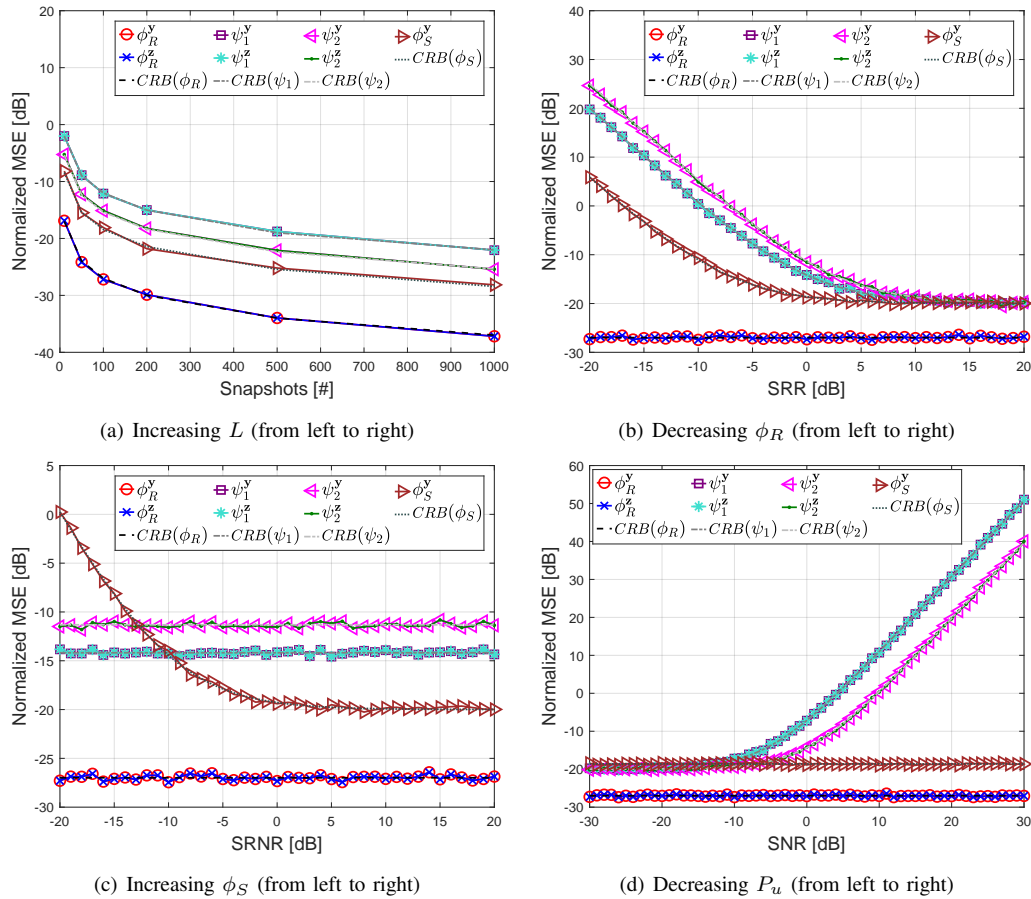(d) Decreasing P_u (from left to right)

Fig. 1: Normalized MSEs and CRBs as a function of: (a) Number of snapshots, (b) SRR, (c) SRNR and (d) SNR. The parameters trends are depicted in the caption of each subfigure.

TABLE II: Nominal Parameters

Definition	Symbol	Nominal value
Microphone number	N	8
Inter-microphone distance	d	6 cm
Direction of arrival	θ	$0°$
Frequency	f	2 kHz
Noise-PSD rank	T	2
Number of Snapshots	L	100
Speech PSD	ϕ_S	0.5
Rev. PSD	ϕ_R	0.5
Noise Power	P_u	0.5

CRB was also computed, in order to demonstrate the theoretical lower bound on the nMSE.

2) Simulation Results: In Fig. 1(a), the nMSEs are presented as a function of the number of snapshots, L. Clearly, the nMSEs of all the estimators decrease as the number of snapshots increases. As expected from the analytical study, it is evident that the non-blocking-based and the blocking-based MLEs yield the same nMSE, for both the reverberation and noise PSDs. Furthermore, for all quantities the nMSEs coincide with the corresponding CRBs. It should be stressed

that in practical applications of speech processing, the PSDs are highly time-varying, and thus will not remain constant over a large number of snapshots. This experiment only serves as a consistency check for the proposed estimators.

We now study the effect of varying the reverberation level. Let the signal-to-reverberation ratio (SRR) be defined as $\text{SRR} \triangleq 10 \log \left(\frac{\phi_S}{\phi_R} \right)$. In this experiment, we change ϕ_R s.t. the resulting SRR ranges between -20 dB and 20 dB. In Fig. 1(b), the nMSEs are presented as a function of SRR. It is evident that the nMSEs of the reverberation PSD estimators are independent of ϕ_R, and equal to $\frac{1}{L} \frac{1}{N-1-T}$, as manifested by (64). On the other hand, the nMSEs of the noise and speech MLEs decrease as the reverberation level decreases, until reaching the limiting value of $\frac{1}{L}$, which is inline with (70) and (73), respectively.

We now examine the effect of changing the speech PSD level. Let the signal-to-reverberation-plus-noise ratio (SRNR) be defined as $\text{SRNR} \triangleq 10 \log \left(\frac{\phi_S}{\phi_R + P_u} \right)$. We set ϕ_S to several values s.t. the SRNR ranges between -20 dB and 20 dB. In Fig. 1(c), the nMSEs are presented as a function of SRNR. It is shown that the speech PSD estimator is improved as the speech level increases. For the reverberation and noise PSD estimators, the performance is independent of ϕ_S. Obviously, the blocking-based estimators are not affected by the value

of the speech PSD. It is further shown that the non-blocking-based reverberation and noise estimators also produce nMSEs that are independent of ϕ_S, as implied by (64) and (68). This behaviour can be explained as follows. Both the reverberation and the noise PSD estimators apply a projection matrix onto the subspace orthogonal to the speech subspace, i.e. eliminate the speech component by generating a null towards its direction, as manifested by (22) and (29), respectively. As a result, they are not affected by any change in the speech level.

The effect of increasing the noise level is now examined. Let the signal-to-noise ratio (SNR) be defined as SNR $\triangleq$ $10\log\left(\frac{\phi_S}{P_u}\right)$. We change P_u s.t. the SNR varies between $-30\,\text{dB}$ and $30\,\text{dB}$. In Fig. 1(d), the nMSEs are presented as a function of SNR. It is evident that the performance of the noise PSD estimators degrades as P_u decreases. In contrast, the nMSEs of the reverberation and speech MLEs are independent of the noise level, as manifested by (64) and (71), respectively. This is due to the fact that the reverberation and speech MLEs eliminate the noise components by directing null towards the noise subspace, as implied by (22) and (25), respectively.

B. Experiments with Recorded Room Impulse Responses

The performance of the proposed PSD estimators is now evaluated in a realistic acoustic scenario, for the task of speech dereverberation and noise reduction. In our experiments, microphone signals were synthesized using real speech signals and measured RIRs. The proposed PSD estimators were used in order to calculate the MCWF.

1) Competing Algorithms: The proposed method is compared to [15], in which the MCWF is implemented using the blocking-based or the non-blocking-based LS estimators. Therein, a spatially homogeneous noise sound field is assumed, namely $\boldsymbol{\Phi}_{\mathbf{u}}(m,k) = \phi_U(m,k)\boldsymbol{\Gamma}_U(k)$, where $\boldsymbol{\Gamma}_U(k)$ is a known time-invariant spatial coherence matrix, and $\phi_U(m,k)$ denotes the unknown time-varying PSD, which has to be estimated, along with the speech and reverberation PSDs. Although this method considers a different noise model, it is chosen as the baseline since this is the only work that estimates also the noise PSD.

2) Implementation of the MCWF: It is well-known that the MCWF can be decomposed into an MVDR beamformer followed by a single-channel Wiener postfilter [22], [23]:

$$\hat{s}_{\text{MCWF}}(m) = \underbrace{\frac{\hat{\gamma}(m)}{\hat{\gamma}(m)+1}}_{H_W(m)}\underbrace{\frac{\mathbf{g}_d^{\text{H}}\hat{\boldsymbol{\Phi}}_i^{-1}(m)}{\mathbf{g}_d^{\text{H}}\hat{\boldsymbol{\Phi}}_i^{-1}(m)\mathbf{g}_d}}_{\mathbf{w}_{\text{MVDR}}^{\text{H}}(m)}\mathbf{y}(m), \qquad (76)$$

where

$$\hat{\gamma}(m) = \frac{\hat{\phi}_S(m)}{\hat{\phi}_{RE}(m)} \qquad (77)$$

denotes the SRNR at the output of the MVDR, and $\hat{\phi}_{RE}(m) \triangleq \left(\mathbf{g}_d^{\text{H}}\hat{\boldsymbol{\Phi}}_i^{-1}(m)\mathbf{g}_d\right)^{-1}$ is the residual interference at the MVDR output.

The implementation of (77) requires the estimate of the speech PSD, which is missing in the blocking-based framework. By substituting the obtained blocking-based reverberation and noise estimates, namely $\phi_R^{\text{ML},\bar{\mathbf{z}}}$ and $\boldsymbol{\Psi}_{\mathbf{v}}^{\text{ML},\bar{\mathbf{z}}}$, into the
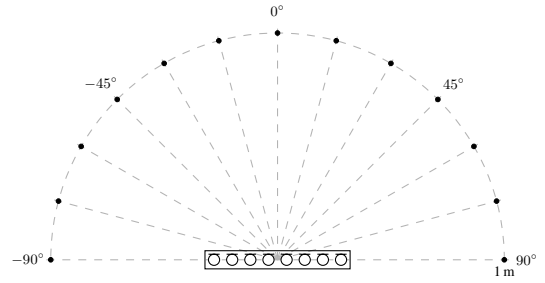

Fig. 2: Geometric setup.

general likelihood function in (18), the maximization becomes a one-dimensional optimization problem, and a closed-form solution is available [9], [24]:

$$\phi_S^{\bar{\mathbf{z}}}(m) = \mathbf{w}_{\text{MVDR}}^{\text{H}}(m)\Big(\mathbf{R_y}(m) - \phi_R^{\text{ML},\bar{\mathbf{z}}}(m)\boldsymbol{\Gamma}_R$$
$$- \mathbf{V}\boldsymbol{\Psi}_{\mathbf{v}}^{\text{ML},\bar{\mathbf{z}}}(m)\mathbf{V}^{\text{H}}\Big)\mathbf{w}_{\text{MVDR}}(m). \qquad (78)$$

However, it was shown in [15] that rather than using (77), better dereverberation performance is obtained by using the decision-directed approach [25], where $\gamma(m)$ is estimated by

$$\hat{\gamma}_{\text{DD}}(m) = \beta\frac{|\hat{s}(m-1)|^2}{\hat{\phi}_{RE}(m-1)} + (1-\beta)\frac{\hat{\phi}_{S_i}(m)}{\hat{\phi}_{RE}(m)}, \qquad (79)$$

where β is a weighting factor, and $\hat{\phi}_{S_i}$ is an instantaneous estimate based on the MVDR output [8]:

$$\hat{\phi}_{S_i}(m) = \max\left(|\mathbf{w}_{\text{MVDR}}^{\text{H}}(m)\mathbf{y}(m)|^2 - \hat{\phi}_{RE}(m), 0\right). \quad (80)$$

In our experiments, the MCWF was implemented with the two variants of computing $\hat{\gamma}$: i) The direct implementation in (77), which will be referred to as Dir; and ii) the decision-directed implementation in (79) with the speech PSD estimated as in (80), denoted henceforth as DD.

3) Performance Measures: The speech quality was evaluated in terms of three common objective measures, namely perceptual evaluation of speech quality (PESQ) [26], frequency-weighted segmental SNR (fwSNRseg) [27] and log-likelihood ratio (LLR) [27]. The measures were computed by comparing $\hat{s}_{\text{MCWF}}(m)$ with the direct speech signal as received by the reference microphone, obtained by filtering the anechoic speech with the direct path. In the sequel, we present the improvement in these performance measures, namely ΔPESQ, ΔfwSNRseg and ΔLLR, computed as the measure difference between the output signal (i.e. $\hat{s}_{\text{MCWF}}$) and the noisy and reverberant signal at the reference microphone (i.e. y_1).

4) Experimental Setup: We used RIRs from the database presented in [28]. The impulse responses were recorded in the $6 \times 6 \times 2.4$ m acoustic lab of the Engineering Faculty at Bar-Ilan University (BIU). The reverberation time was set to $T_{60} \in \{360, 610\}$ msec. The RIRs were recorded by a ULA of $N = 8$ microphones with inter-distances of 8 cm. A loudspeaker was positioned at a distance of 1 m from the array, at angles $\{-90°, -75°, -60°, \ldots, 90°\}$, as illustrated in Fig. 2. The performance is evaluated on 10 experiments with different speaker angles, which were randomly selected from the given set of angles. The 10 chosen RIRs were convolved

TABLE III: Speech Quality Measures, $T_{60} = 360$ msec

Alg.\RSNR	ΔPESQ				ΔfwSNRseg [dB]				ΔLLR [dB]			
	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
Blocking LS [15] Dir	0.56	0.70	0.77	0.76	11.85	11.02	9.80	8.67	-0.55	-0.52	-0.45	-0.37
Blocking LS [15] DD	0.60	0.75	0.85	0.88	13.05	11.67	10.23	8.88	-0.50	-0.45	-0.36	-0.27
Blocking ML Dir	0.74	0.88	0.93	0.90	12.56	11.77	10.61	9.39	**-0.74**	**-0.65**	**-0.52**	**-0.40**
Blocking ML DD	0.79	0.93	0.99	0.99	15.57	13.16	10.75	8.66	-0.67	-0.50	-0.30	-0.12
Non-blocking LS [15] Dir	0.70	0.81	0.85	0.81	13.52	11.69	9.93	8.57	-0.54	-0.49	-0.41	-0.33
Non-blocking LS [15] DD	0.67	0.79	0.89	0.89	13.56	11.82	10.03	8.57	-0.55	-0.48	-0.38	-0.28
Non-blocking ML Dir	**0.98**	**1.12**	**1.15**	**1.12**	**15.59**	**13.69**	**11.67**	**9.77**	**-0.74**	-0.61	-0.46	-0.32
Non-blocking ML DD	0.79	0.93	0.99	0.99	15.57	13.16	10.75	8.66	-0.67	-0.50	-0.30	-0.12

TABLE IV: Speech Quality Measures, $T_{60} = 610$ msec

Alg.\RSNR	ΔPESQ				ΔfwSNRseg [dB]				ΔLLR [dB]			
	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB	0dB	5dB	10dB	15dB
Blocking LS [15] Dir	0.42	0.50	0.54	0.56	10.78	10.13	9.20	8.38	-0.51	-0.49	-0.44	-0.37
Blocking LS [15] DD	0.46	0.54	0.62	0.67	12.49	11.54	10.46	9.47	-0.45	-0.41	-0.36	-0.30
Blocking ML Dir	0.51	0.59	0.63	0.63	10.10	9.57	8.68	7.95	-0.58	**-0.53**	**-0.46**	**-0.38**
Blocking ML DD	0.61	0.71	0.78	0.80	**14.93**	**13.10**	**11.30**	**9.67**	-0.57	-0.44	-0.28	-0.13
Non-blocking LS [15] Dir	0.56	0.60	0.63	0.62	13.22	11.49	9.93	8.70	-0.51	-0.48	-0.42	-0.36
Non-blocking LS [15] DD	0.51	0.59	0.67	0.71	13.37	12.07	10.66	8.73	-0.51	-0.47	-0.40	-0.32
Non-blocking ML Dir	**0.69**	**0.78**	**0.80**	**0.81**	13.88	12.39	10.81	9.47	**-0.62**	-0.52	-0.41	-0.30
Non-blocking ML DD	0.61	0.71	0.78	0.80	**14.93**	**13.10**	**11.30**	**9.67**	-0.57	-0.44	-0.28	-0.13

with 3-5 sec long utterances of 5 male and 5 female speakers that were drawn from the TIMIT database [29].

For the additive noise, we used *non-stationary* directional noise signals. Two non-stationary noise sources (with time-varying speech-like spectrum) were positioned at a distance of 1 m from the array. In each experiment, the angles of the noise sources were randomly selected, avoiding the angle of the speaker. The microphone signals were synthesized by adding the noises to the reverberant speech signals with several reverberant signal-to-noise ratio (RSNR) levels. Finally, a 1 sec noise segment was preceded to the reverberant speech signal. Based on this segment, a sample noise covariance matrix is computed, and then the EVD is applied in order to extract the spatial basis $\mathbf{V}$ that replaces the noise ATF matrix $\mathbf{A_u}$ (see (14)).

The following values of the parameters were used. The sampling rate was set to 16 KHz, the STFT was computed with windows of 32 msec and 75% overlap between adjacent time frames. As the experiment consists of real-life non-stationary speech signals, the sample covariance matrix $\mathbf{R_y}(m)$ was estimated using recursive averaging [9] with a smoothing factor of $\alpha = 0.7$, rather than the moving-window averaging of (19). The same applies for $\mathbf{R_z}(m)$. The smoothing parameter for the decision-directed in (79) was set to $\beta = 0.9$. The gain of the single channel postfilter was lower bounded by -15 dB.

5) Experimental Results: First, we examine the performance obtained with various RSNR levels. Tables III and IV summarize the results for both reverberation levels. Note that negative ΔLLR indicates a performance improvement. The best results are highlighted in boldface. It is evident that the

proposed ML methods outperform the baseline methods of [15] for all the considered scenarios. The advantage of the proposed method can be attributed to the fact that the full $T \times T$ spatial noise PSD matrix is estimated in each frame, and thus the non-stationarity of all the noise sources can be tracked, whereas the baseline model of a time-invariant spatial coherence matrix with a single controllable gain parameter cannot track this dynamics.

In order to analyze the behaviour of the various ML implementations, recall that the non-blocking-based and the blocking-based noise and reverberation MLEs were proved to be identical in Section III-C. Since both the non-blocking ML DD and the blocking ML DD use also the same speech PSD estimate in (80), it follows that both methods yield the same performance. However, the non-blocking ML Dir and the blocking ML Dir implementations differ, due to the fact that they use different speech PSD estimators in (24) and (78), respectively.

From Tables III and IV, it is evident that the non-blocking-based ML Dir method is preferable for its high and stable scores, as it obtains the best ΔPESQ and ΔfwSNRseg scores in most cases, and also yields high ΔLLR results. The advantage of the non-blocking ML Dir over the blocking ML Dir can be attributed to the fact that the non-blocking ML Dir uses the speech PSD estimate in (24), which is independent of the estimated noise PSD. In contrast, the blocking-based ML Dir estimates the speech PSD with (78), which depends on the estimated noise PSD, and thus can be affected by estimation errors of the noise PSD. Indeed, there is a slight disadvantage for the non-blocking ML Dir in terms of ΔLLR, which mainly
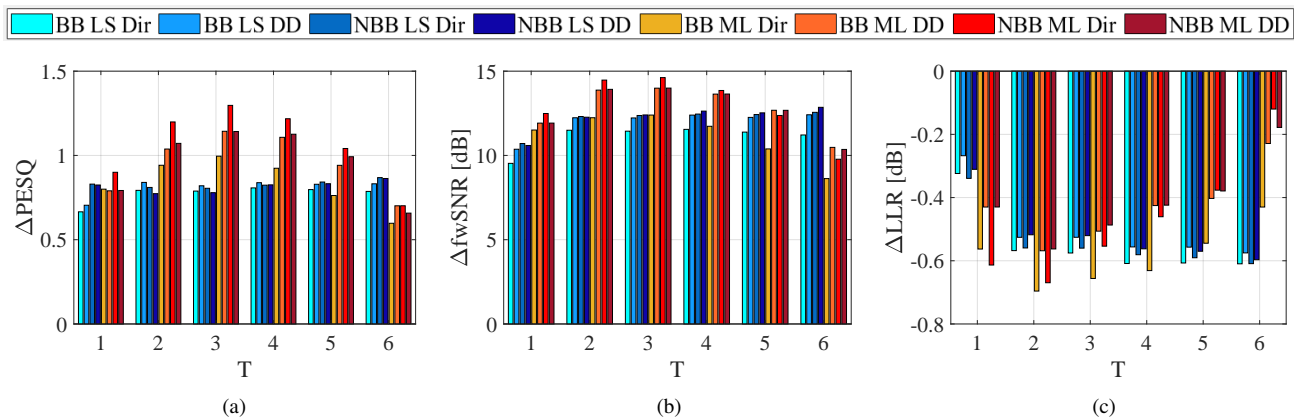
Fig. 3: Results versus T, for $T_{60} = 360$ msec, $N = 8$, RSNR = 5dB: (a) ΔPESQ, (b) ΔfwSNR, and (c) ΔLLR.
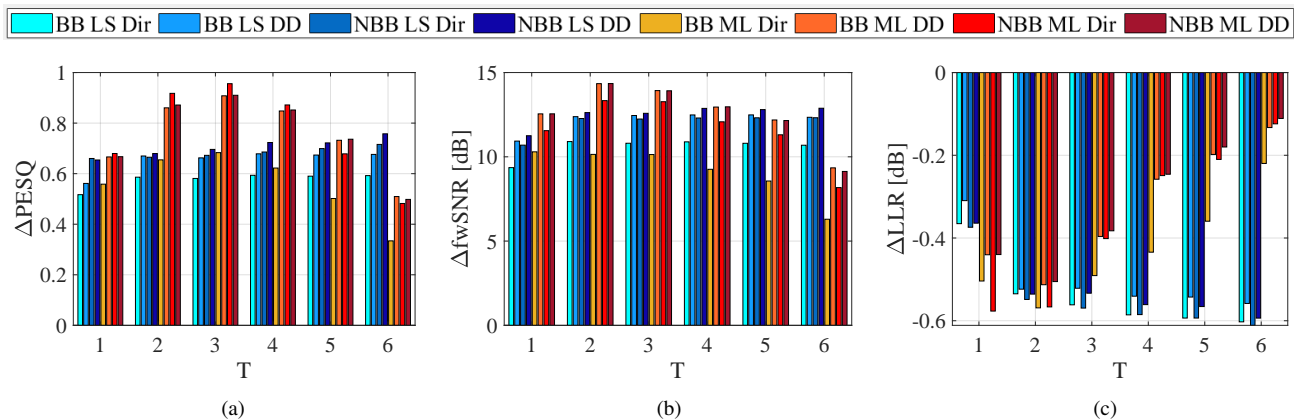
Fig. 4: Results versus T, for $T_{60} = 610$ msec, $N = 8$, RSNR = 5dB: (a) ΔPESQ, (b) ΔfwSNR, and (c) ΔLLR.

quantifies the performance due to speech distortion. This can be related to the fact that the speech PSD estimate in (24) generates nulls towards the noise sources, which in some cases can affect speech distortion, for example when the speech and noise sources are not fully spatially separated.

Comparing the direct and the DD implementations of the non-blocking ML approach, it can be observed that there is a clear advantage for the direct implementation in most cases. This can be explained by the fact that the direct method optimally estimates the speech PSD by maximizing the ML criterion, while the decision-directed approach estimates the speech PSD using (79)–(80), which is derived in a more heuristic manner. However, in adverse conditions of high noise and reverberation levels, the DD has some advantage due to its use of a smoother estimator compared to the instantaneous estimator of the direct form, thus improving the robustness of the speech PSD estimate. To conclude, the non-blocking ML Dir is the preferable method among the proposed estimators.

Next, the performance is investigated as a function of the number of noise sources. A representative scenario with RSNR = 5 dB and $N = 8$ was inspected. Since our model assumes that $T \leq N - 2$, the following values of T were examined: $T \in \{1, \ldots, 6\}$. Figs. 3 and 4 depict the results for both reverberation levels, where NBB denotes the non-blocking-based method, and BB refers to the blocking-based

method. It is shown that the proposed methods outperform the baseline methods [15] in most cases, which is inline with the results obtained in Tables III and IV, showing the advantage of the proposed rank-deficient noise model. The baseline methods have an advantage only for values of T close to $N-2$. This can be attributed to the fact that when the number of non-stationary noise sources increases, their sum tends to be stationary, especially in high reverberation conditions. In this case, it might be preferred to use the simpler approximation of the noise PSD matrix, namely a time-invariant spatial coherence matrix multiplied by a scalar time-varying PSD, rather than the proposed model that tracks the non-stationarity of each of the noises, and requires the estimation of a larger amount of parameters.

Finally, we test the influence of the number of microphones on the performance. Figs. 5 and 6 depict the measures obtained with different number of microphones, i.e. $N \in \{4, 6, 8\}$, for both reverberation times, where $T = 2$ and RSNR = 5 dB. It is evident that the proposed methods outperform the baseline methods in almost all cases.

Fig. 7 depicts several sonogram examples of the various signals at $T_{60} = 610$ msec and RSNR of 15 dB. Fig. 7(a) shows s, the direct speech signal as received by the first microphone. Fig. 7(b) depicts y_1, the noisy and reverberant signal at the first microphone. Figs. 7(c) and 7(d) show
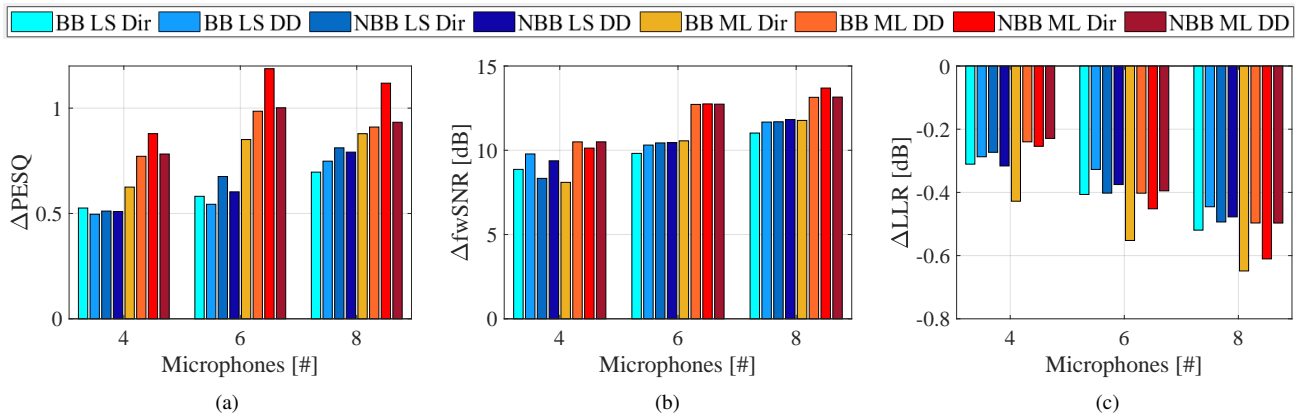
Fig. 5: Results versus N, for $T_{60} = 360$ msec, $T = 2$, RSNR = 5dB: (a) ΔPESQ, (b) ΔfwSNR, and (c) ΔLLR.
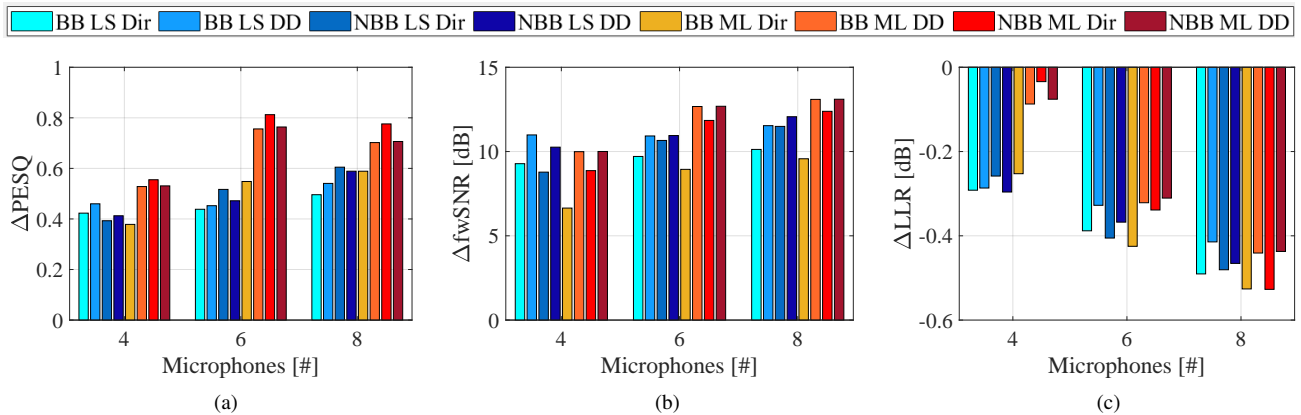

Fig. 6: Results versus N, for $T_{60} = 610$ msec, $T = 2$, RSNR = 5dB: (a) ΔPESQ, (b) ΔfwSNR, and (c) ΔLLR.

the MCWF output computed with (77), using the proposed blocking-based and non-blocking-based MLEs, respectively. We conclude that the application of the MCWF, implemented based on the proposed MLEs, reduces significantly noise and reverberation, while maintaining low speech distortion.

VI. CONCLUSIONS

In this contribution, we discussed the problem of joint dereverberation and noise reduction, in the presence of directional noise sources, forming a rank-deficient noise PSD matrix. As opposed to state-of-the-art methods which assume the knowledge of the noise PSD matrix, we propose to estimate also the time-varying noise PSD matrix, assuming that a basis that spans the noise ATF subspace is known. MLEs of the reverberation, speech and noise PSDs are derived for both the non-blocking-based and the blocking-based methods. The resulting MLEs are of closed-form and thus have low computational complexity. The proposed estimators are theoretically analyzed and compared. For both the reverberation and the noise PSD estimators, it is shown that the non-blocking-based MLE and the blocking-based MLE are identical. The estimators were shown to be unbiased, and the corresponding MSEs were calculated. Moreover, CRBs on the various PSDs were derived, and were shown to be identical to the MSEs of the proposed estimators. The discussion is supported by an

experimental study based on both simulated data and real-life audio signals. It is shown that using the proposed estimators yields a large performance improvement with respect to a competing method.

APPENDIX A

The proof follows the lines of [24]. First, we use (15) and rewrite the PSD matrix of the microphone signals in (9) as

$$\mathbf{\Phi}_{\mathbf{y}}(m) = \mathbf{A}\mathbf{\Phi}_{sv}(m)\mathbf{A}^{\mathrm{H}} + \phi_R(m)\,\mathbf{\Gamma}_R, \qquad (81)$$

where $\mathbf{A}$ is defined in (23) and $\mathbf{\Phi}_{sv}(m) \in \mathbb{C}^{(T+1)\times(T+1)}$ is given by

$$\mathbf{\Phi}_{sv}(m) = \left[\begin{array}{c|c} \phi_S(m) & \mathbf{0}^{\mathrm{H}} \\ \hline \mathbf{0} & \mathbf{\Psi}_{\mathbf{v}}(m) \end{array}\right]. \qquad (82)$$

Then, we define $\mathbf{P} \in \mathbb{C}^{N\times N}$ as an orthogonal projection matrix onto the speech-plus-noise subspace

$$\mathbf{P} = \mathbf{A}\left(\mathbf{A}^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{H}}\mathbf{\Gamma}_R^{-1}, \qquad (83)$$

(a) Clean direct speech at microphone #1.

(b) Noisy and Reverberant signal at microphone #1.

(c) Output of the MCWF (direct implementation) using the proposed blocking-based MLEs.

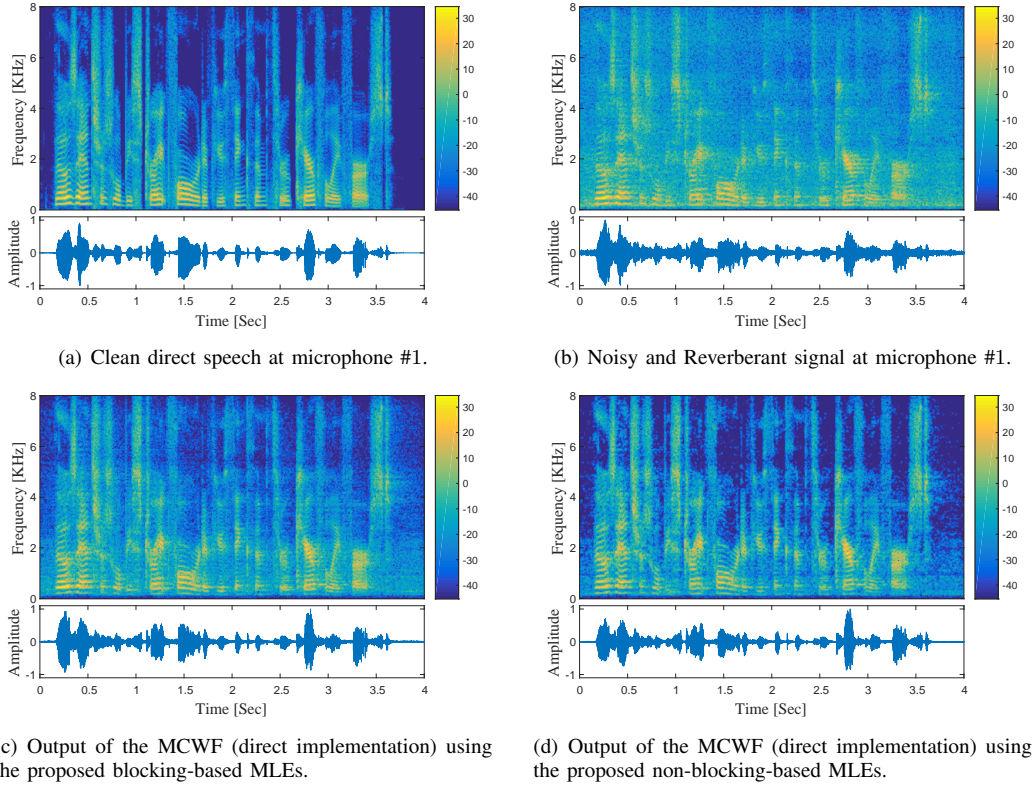(d) Output of the MCWF (direct implementation) using the proposed non-blocking-based MLEs.

Fig. 7: An example of audio sonograms, with $T_{60} = 610$ msec and RSNR = 15dB.

and $\mathbf{Q} = \mathbf{I}_N - \mathbf{P}$ is the orthogonal complement. We will make use of the following properties, which can be easily verified:

$$\mathbf{P} = \mathbf{PP}, \tag{84a}$$

$$\mathbf{Q} = \mathbf{QQ}, \tag{84b}$$

$$\mathbf{Q}\boldsymbol{\Phi}_{\mathbf{y}}(m) = \phi_R(m)\mathbf{Q}\boldsymbol{\Gamma}_R, \tag{84c}$$

$$\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{Q} = \phi_R^{-1}(m)\boldsymbol{\Gamma}_R^{-1}\mathbf{Q}, \tag{84d}$$

$$\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{A} = \boldsymbol{\Gamma}_R^{-1}\mathbf{A}\left(\boldsymbol{\Phi}_{sv}(m)\mathbf{A}^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{A} + \phi_R(m)\mathbf{I}\right)^{-1}. \tag{84e}$$

Using [11, Eq. (15.47-15.48)], the derivative of the log likelihood function w.r.t. $\phi_R(m)$ is given by

$$\mathcal{L}\big(\phi_R(m)\big) \triangleq \frac{\partial \log p\big(\bar{\mathbf{y}}(m); \boldsymbol{\phi}(m)\big)}{\partial \phi_R(m)}$$
$$= L\,\mathrm{Tr}\left[\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{R}_{\mathbf{y}}(m) - \mathbf{I}\right)\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\boldsymbol{\Gamma}_R\right]. \tag{85}$$

Using (84a) and (84b) it follows that

$$\mathcal{L}\big(\phi_R(m)\big) = L\,\mathrm{Tr}\Big[\mathbf{Q}\boldsymbol{\Gamma}_R\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{R}_{\mathbf{y}}(m) - \mathbf{I}\right)\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{Q}$$
$$+ \mathbf{P}\boldsymbol{\Gamma}_R\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{R}_{\mathbf{y}}(m) - \mathbf{I}\right)\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{P}\Big]. \tag{86}$$

However, by substituting (83) into (86), it can be shown that the second term vanishes (follows from setting the derivative

of the likelihood w.r.t. $\boldsymbol{\Phi}_{sv}$ in (88) to zero). Hence,

$$\mathcal{L}\big(\phi_R(m)\big) = L\,\mathrm{Tr}\Big[\mathbf{Q}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\big(\mathbf{R}_{\mathbf{y}}(m) - \boldsymbol{\Phi}_{\mathbf{y}}(m)\big)$$
$$\times \boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{Q}\Big]$$
$$\overset{(a)}{=} L\phi_R^{-2}(m)\mathrm{Tr}\left[\mathbf{Q}\big(\mathbf{R}_{\mathbf{y}}(m) - \boldsymbol{\Phi}_{\mathbf{y}}(m)\big)\boldsymbol{\Gamma}_R^{-1}\mathbf{Q}\right]$$
$$\overset{(b)}{=} L\phi_R^{-2}(m)\mathrm{Tr}\left[\mathbf{Q}\mathbf{R}_{\mathbf{y}}(m)\boldsymbol{\Gamma}_R^{-1} - \phi_R(m)\mathbf{Q}\right], \tag{87}$$

where (a) follows from (84c) and (84d), and (b) follows from (84b) and (84c). Finally, using (22) we have $\mathrm{Tr}\left[\mathbf{Q}\right] = N - (T+1)$. Thus, setting (87) to zero yields (21).

Appendix B

Using [24, Eq. (12)], the derivative of the log likelihood function $\log p\big(\bar{\mathbf{y}}(m); \boldsymbol{\phi}(m)\big)$ w.r.t. $\boldsymbol{\Phi}_{sv}$ is given by

$$\mathcal{L}\big(\boldsymbol{\Phi}_{sv}(m)\big) \triangleq \frac{\partial \log p\big(\bar{\mathbf{y}}(m); \boldsymbol{\phi}(m)\big)}{\partial \boldsymbol{\Phi}_{sv}(m)}$$
$$= L\left(\mathbf{A}^{\mathsf{H}}\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{R}_{\mathbf{y}}(m) - \mathbf{I}\right)\boldsymbol{\Phi}_{\mathbf{y}}^{-1}(m)\mathbf{A}\right)^{\top}. \tag{88}$$

Using (84e) and setting the result to zero yields

$$\boldsymbol{\Phi}_{sv}^{\mathrm{ML}}(m) = \left(\mathbf{A}^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\left(\mathbf{R}_{\mathbf{y}}(m) - \phi_R^{\mathrm{ML},\bar{\mathbf{y}}}(m)\boldsymbol{\Gamma}_R\right)$$
$$\times \boldsymbol{\Gamma}_R^{-1}\mathbf{A}\left(\mathbf{A}^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{A}\right)^{-1}. \tag{89}$$

In order to simplify the expression of (89), we define a partitioned matrix

$$\mathbf{A}^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{A} = \left[\begin{array}{c|c} \mathbf{g}_d^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{g}_d & \mathbf{g}_d^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{V} \\ \hline \mathbf{V}^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{g}_d & \mathbf{V}^{\mathsf{H}}\boldsymbol{\Gamma}_R^{-1}\mathbf{V} \end{array}\right]. \tag{90}$$

Using the formula of the inverse of a partitioned matrix and then taking the $(1,1)$ entry of $\mathbf{\Phi}_{sv}^{\mathrm{ML}}$, the MLE of ϕ_S in (24) is obtained.

APPENDIX C

We note that $\mathbf{\Psi}_{\mathbf{v}}(m)$ is the $T \times T$ lower-right block of the full $(T+1) \times (T+1)$ matrix $\mathbf{\Phi}_{sv}(m)$ in (82). Using again the formula for the inverse of the partitioned matrix in (90) and taking the corresponding entries, yields (28).

APPENDIX D

The proof follows by substituting the inverse of the partitioned matrix (90) into (22), and then using the definitions in (25) and (29).

APPENDIX E

The PSD matrix of the BM output is given in (42) as

$$\mathbf{\Phi}_{\mathbf{z}}(m) = \widetilde{\mathbf{V}} \mathbf{\Psi}_{\mathbf{v}}(m) \widetilde{\mathbf{V}}^{\mathrm{H}} + \phi_R(m) \mathbf{B}^{\mathrm{H}} \mathbf{\Gamma}_R \mathbf{B}, \qquad (91)$$

where $\widetilde{\mathbf{V}}$ is defined in (41). The proof is now similar to that of Appendix A, with the following changes: $\mathbf{A}$, $\mathbf{\Phi}_{sv}(m)$, $\mathbf{\Gamma}_R$ and $\mathbf{R}_{\mathbf{y}}(m)$ are replaced with $\widetilde{\mathbf{V}}$, $\mathbf{\Psi}_{\mathbf{v}}(m)$, $\mathbf{B}^{\mathrm{H}} \mathbf{\Gamma}_R \mathbf{B}$ and $\mathbf{R}_{\mathbf{z}}(m)$, respectively.

APPENDIX F

Using (42), the MLE of $\mathbf{\Psi}_{\mathbf{v}}(m)$ can be calculated in a similar manner to (89), with

$$\mathbf{\Psi}_{\mathbf{v}}^{\mathrm{ML},\bar{\mathbf{z}}}(m) = \widetilde{\mathbf{W}}_{\mathbf{u}}^{\mathrm{H}} \left(\mathbf{R}_{\mathbf{z}}(m) - \phi_R^{\mathrm{ML},\bar{\mathbf{z}}}(m) \left(\mathbf{B}^{\mathrm{H}} \mathbf{\Gamma}_R \mathbf{B} \right) \right) \widetilde{\mathbf{W}}_{\mathbf{u}}, \qquad (92)$$

where $\widetilde{\mathbf{W}}_{\mathbf{u}}$ is given in (49).

APPENDIX G

Substituting (41) into (49), and then using (54), leads to

$$\widetilde{\mathbf{W}}_{\mathbf{u}}^{\mathrm{H}} = \left(\mathbf{V}^{\mathrm{H}} \mathbf{P}_g^{\perp} \mathbf{\Gamma}_R^{-1} \mathbf{V} \right)^{-1} \mathbf{V}^{\mathrm{H}} \mathbf{B} \left(\mathbf{B}^{\mathrm{H}} \mathbf{\Gamma}_R \mathbf{B} \right)^{-1}. \qquad (93)$$

Right multiplying (93) by $\mathbf{B}^{\mathrm{H}}$, using again (54) and then comparing to (29), yields (52).

APPENDIX H

Using (19), (21) and the i.i.d. assumption, the variance of (65) can be recast as

$$\mathrm{var}\left(\psi_i^{\mathrm{ML},\bar{\mathbf{y}}}(m) \right)$$
$$= \frac{1}{L} \mathrm{var}\left[\mathbf{y}^{\mathrm{H}}(m) \left(\mathbf{w}_i \mathbf{w}_i^{\mathrm{H}} - \frac{\mathbf{w}_i^{\mathrm{H}} \mathbf{\Gamma}_R \mathbf{w}_i}{N-1-T} \mathbf{\Gamma}_R^{-1} \mathbf{Q} \right) \mathbf{y}(m) \right]. \quad (94)$$

To proceed, we use (62), (59), and (33) to obtain

$$\mathrm{var}\left(\psi_i^{\mathrm{ML},\bar{\mathbf{y}}}(m) \right)$$
$$= \frac{1}{L} \left[\left(\mathbf{w}_i^{\mathrm{H}} \mathbf{\Phi}_{\mathbf{y}}(m) \mathbf{w}_i \right)^2 + \frac{\left(\mathbf{w}_i^{\mathrm{H}} \phi_R(m) \mathbf{\Gamma}_R \mathbf{w}_i \right)^2}{N-1-T} \right]. \quad (95)$$

Finally, using (9) and (31) yields (68). The right hand side of (69) is obtained by substituting (66) into the definition of ξ_i, and noting that $\left(\mathbf{P}_{\widetilde{\mathbf{V}}_i}^{\perp} \mathbf{P}_g^{\perp} \right)^{\mathrm{H}} = \mathbf{\Gamma}_R \mathbf{P}_g^{\perp} \mathbf{P}_{\widetilde{\mathbf{V}}_i}^{\perp} \mathbf{P}_g^{\perp} \mathbf{\Gamma}_R^{-1}$.

APPENDIX I

In this Appendix, the CRB on the reverberation PSD is derived. We will make use of the following identities [30]:

$$\mathrm{Tr}\left[\mathbf{X}\mathbf{Y} \right] = \left(\mathrm{vec}\left(\mathbf{X}^{\mathrm{H}} \right) \right)^{\mathrm{H}} \mathrm{vec}\left(\mathbf{Y} \right), \qquad (96)$$

$$\mathrm{vec}\left(\mathbf{X}\mathbf{Y}\mathbf{Z} \right) = \left(\mathbf{Z}^{\top} \otimes \mathbf{X} \right) \mathrm{vec}\left(\mathbf{Y} \right), \qquad (97)$$

$$\left(\mathbf{W} \otimes \mathbf{X} \right)\left(\mathbf{Y} \otimes \mathbf{Z} \right) = \left(\mathbf{W}\mathbf{Y} \right) \otimes \left(\mathbf{X}\mathbf{Z} \right). \qquad (98)$$

Using the definitions of $\mathbf{\Phi}_{\mathbf{y}}$ and $\mathbf{\Phi}_{sv}$ in (81) and (82), we denote the set of unknown parameters by $\boldsymbol{\alpha} \triangleq \left[\phi_R, \phi_{SV}^{\top} \right]^{\top} \in \mathbb{C}^{\left((T+1)^2+1 \right)}$, where $\phi_{SV} \triangleq \mathrm{vec}\left(\mathbf{\Phi}_{sv} \right) \in \mathbb{C}^{(T+1)^2}$. As $\mathbf{\Phi}_{\mathbf{y}}$ is a PSD matrix of a Gaussian vector, the Fisher information of each pair of parameters is given by [31]:

$$\mathrm{I}_{ij}^{\bar{\mathbf{y}}} = L \, \mathrm{Tr}\left[\mathbf{\Phi}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{\Phi}_{\mathbf{y}}}{\partial \alpha_i} \mathbf{\Phi}_{\mathbf{y}}^{-1} \frac{\partial \mathbf{\Phi}_{\mathbf{y}}}{\partial \alpha_j} \right], \qquad (99)$$

where $\mathrm{I}_{ij}^{\bar{\mathbf{y}}}$ is the Fisher information of α_i and α_j and $i,j = 1, \ldots, (T+1)^2 + 1$. In order to facilitate the derivation, we use (97) and vectorize (81) to obtain an $N^2 \times 1$ vector:

$$\phi_y \triangleq \mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{y}} \right) = \left(\mathbf{A}^* \otimes \mathbf{A} \right) \mathrm{vec}\left(\mathbf{\Phi}_{sv} \right) + \phi_R \mathrm{vec}\left(\mathbf{\Gamma}_R \right). \quad (100)$$

Using (96), (97), (99) and (100), the full Fisher information matrix (FIM) writes [32], [33]

$$\frac{1}{L} \mathrm{I}^{\bar{\mathbf{y}}} = \left(\frac{\partial \phi_y}{\partial \boldsymbol{\alpha}^{\top}} \right)^{\mathrm{H}} \left(\mathbf{\Phi}_{\mathbf{y}}^{-\top} \otimes \mathbf{\Phi}_{\mathbf{y}}^{-1} \right) \left(\frac{\partial \phi_y}{\partial \boldsymbol{\alpha}^{\top}} \right). \qquad (101)$$

Next, we define the following partitioned matrix:

$$[\mathbf{g}|\mathbf{\Delta}] \triangleq \left(\mathbf{\Phi}_{\mathbf{y}}^{-\top/2} \otimes \mathbf{\Phi}_{\mathbf{y}}^{-1/2} \right) \left[\frac{\partial \phi_y}{\partial \phi_R} \middle| \frac{\partial \phi_y}{\partial \phi_{SV}^{\top}} \right]. \qquad (102)$$

Using (102) and (98), the FIM in (101) writes

$$\frac{1}{L} \mathrm{I}^{\bar{\mathbf{y}}} = \left[\frac{\mathbf{g}^{\mathrm{H}}}{\mathbf{\Delta}^{\mathrm{H}}} \right] [\mathbf{g}|\mathbf{\Delta}] = \left[\begin{array}{c|c} \mathbf{g}^{\mathrm{H}}\mathbf{g} & \mathbf{g}^{\mathrm{H}}\mathbf{\Delta} \\ \hline \mathbf{\Delta}^{\mathrm{H}}\mathbf{g} & \mathbf{\Delta}^{\mathrm{H}}\mathbf{\Delta} \end{array} \right]. \qquad (103)$$

The CRB for ϕ_R is given by $\mathrm{CRB}\left(\phi_R \right) = \left[\left(\mathrm{I}^{\bar{\mathbf{y}}} \right)^{-1} \right]_{1,1}$. Using the formula of the inverse of a partitioned matrix, the CRB writes

$$\mathrm{CRB}\left(\phi_R \right) = \frac{1}{L} \left(\mathbf{g}^{\mathrm{H}}\mathbf{g} - \mathbf{g}^{\mathrm{H}}\mathbf{\Delta} \left(\mathbf{\Delta}^{\mathrm{H}}\mathbf{\Delta} \right)^{-1} \mathbf{\Delta}^{\mathrm{H}}\mathbf{g} \right)^{-1}$$
$$= \frac{1}{L} \left(\mathbf{g}^{\mathrm{H}} \mathbf{\Pi}_{\mathbf{\Delta}}^{\perp} \mathbf{g} \right)^{-1}, \qquad (104)$$

where $\mathbf{\Pi}_{\mathbf{\Delta}}^{\perp} \in \mathbb{C}^{N^2 \times N^2}$ is a projection matrix onto the subspace orthogonal to $\mathbf{\Delta}$:

$$\mathbf{\Pi}_{\mathbf{\Delta}} \triangleq \mathbf{\Delta} \left(\mathbf{\Delta}^{\mathrm{H}}\mathbf{\Delta} \right)^{-1} \mathbf{\Delta}^{\mathrm{H}} , \;\; \mathbf{\Pi}_{\mathbf{\Delta}}^{\perp} = \mathbf{I}_{N^2} - \mathbf{\Pi}_{\mathbf{\Delta}}. \qquad (105)$$

We now simplify (104). First, we use (100) along with (97) to write $\mathbf{g}$ as

$$\mathbf{g} = \left(\mathbf{\Phi}_{\mathbf{y}}^{-\top/2} \otimes \mathbf{\Phi}_{\mathbf{y}}^{-1/2} \right) \mathrm{vec}\left(\mathbf{\Gamma}_R \right)$$
$$= \mathrm{vec}\left(\mathbf{\Phi}_{\mathbf{y}}^{-1/2} \mathbf{\Gamma}_R \mathbf{\Phi}_{\mathbf{y}}^{-1/2} \right). \qquad (106)$$

Similarly, $\boldsymbol{\Delta}$ is simplified by using (100) along with (98),

$$
\begin{aligned}
\boldsymbol{\Delta} &= \left(\boldsymbol{\Phi}_{\mathbf{y}}^{-\top/2} \otimes \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right)\left(\mathbf{A}^* \otimes \mathbf{A}\right) \\
&= \left(\boldsymbol{\Phi}_{\mathbf{y}}^{-\top/2}\mathbf{A}^*\right) \otimes \left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}\right). \quad (107)
\end{aligned}
$$

In order to calculate $\boldsymbol{\Pi}_{\boldsymbol{\Delta}}^{\perp}\mathbf{g}$, the following identity is used [33]:

$$
\boldsymbol{\Pi}_{\mathbf{X}\otimes\mathbf{Y}}^{\perp} = \mathbf{I} \otimes \boldsymbol{\Pi}_{\mathbf{Y}}^{\perp} + \boldsymbol{\Pi}_{\mathbf{X}}^{\perp} \otimes \mathbf{I} - \boldsymbol{\Pi}_{\mathbf{X}}^{\perp} \otimes \boldsymbol{\Pi}_{\mathbf{Y}}^{\perp} \quad (108)
$$

Hence,

$$
\begin{aligned}
\boldsymbol{\Pi}_{\boldsymbol{\Delta}}^{\perp}\mathbf{g} &\overset{(a)}{=} \mathbf{Q}_{\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}\right)^* \otimes \left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}\right)} \mathrm{vec}\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right) \\
&\overset{(b)}{=} \left(\mathbf{I} \otimes \boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp} + \boldsymbol{\Pi}_{\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}\right)^*}^{\perp} \otimes \mathbf{I} \right. \\
&\quad \left. - \boldsymbol{\Pi}_{\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}\right)^*}^{\perp} \otimes \boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\right) \times \mathrm{vec}\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right) \\
&\overset{(c)}{=} \mathrm{vec}\left(\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right. \\
&\quad + \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp} \\
&\quad \left. - \boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right)\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\right), \quad (109)
\end{aligned}
$$

where (a) follows by (106) and (107), (b) follows by (108) and (c) follows by (97). Left multiplying (109) by $\mathbf{g}^{\mathrm{H}}$ yields the reciprocal of the CRB in (104):

$$
\begin{aligned}
\mathbf{g}^{\mathrm{H}}\boldsymbol{\Pi}_{\boldsymbol{\Delta}}^{\perp}\mathbf{g} &\overset{(a)}{=} \left(\mathrm{vec}\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right)\right)^{\mathrm{H}} \\
&\times \mathrm{vec}\left(\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2} + \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\right. \\
&\quad \left. - \boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\left(\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right)\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\right) \\
&\overset{(b)}{=} \mathrm{Tr}\left[2\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right. \\
&\quad \left. - \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Gamma}_R\right], \quad (110)
\end{aligned}
$$

where (a) follows by (106) and (b) follows by (96). To proceed, let us define $\mathbf{D} \in \mathbb{C}^{N \times (N-T-1)}$ as a matrix that spans the nullspace of the matrix $\mathbf{A}$ s.t. $\mathbf{D}^{\mathrm{H}}\mathbf{A} = \mathbf{0}$, i.e. a blocking matrix of the speech-plus-noise signals. Since $\left(\boldsymbol{\Phi}_{\mathbf{y}}^{1/2}\mathbf{D}\right)^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A} = \mathbf{0}$, it follows that [33]

$$
\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp} = \boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{1/2}\mathbf{D}}, \quad (111)
$$

and thus

$$
\begin{aligned}
\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2} &= \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{1/2}\mathbf{D}}\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2} \\
&\overset{(a)}{=} \mathbf{D}\left(\mathbf{D}^{\mathrm{H}}\phi_R\boldsymbol{\Gamma}_R\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{H}} \\
&= \phi_R^{-1}\boldsymbol{\Gamma}_R^{-1}\left(\boldsymbol{\Gamma}_R\mathbf{D}\left(\mathbf{D}^{\mathrm{H}}\boldsymbol{\Gamma}_R\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{H}}\boldsymbol{\Gamma}_R\right)\boldsymbol{\Gamma}_R^{-1} \\
&\overset{(b)}{=} \phi_R^{-2}\boldsymbol{\Gamma}_R^{-1}\left(\boldsymbol{\Phi}_{\mathbf{y}}\mathbf{D}\left(\mathbf{D}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{y}}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{y}}\right)\boldsymbol{\Gamma}_R^{-1} \\
&\overset{(c)}{=} \phi_R^{-2}\boldsymbol{\Gamma}_R^{-1}\left(\boldsymbol{\Phi}_{\mathbf{y}}^{1/2}\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\boldsymbol{\Phi}_{\mathbf{y}}^{1/2}\right)\boldsymbol{\Gamma}_R^{-1}, \quad (112)
\end{aligned}
$$

where (a) and (b) follows since $\mathbf{D}^{\mathrm{H}}\boldsymbol{\Phi}_{\mathbf{y}} = \mathbf{D}^{\mathrm{H}}\phi_R\boldsymbol{\Gamma}_R$ and (c) follows by (111). Substituting (112) into (110) and using the

property $\boldsymbol{\Pi}_{\mathbf{X}}^{\perp}\boldsymbol{\Pi}_{\mathbf{X}}^{\perp} = \boldsymbol{\Pi}_{\mathbf{X}}^{\perp}$, yields

$$
\mathbf{g}^{\mathrm{H}}\boldsymbol{\Pi}_{\boldsymbol{\Delta}}^{\perp}\mathbf{g} = \phi_R^{-2}\mathrm{Tr}\left[\boldsymbol{\Pi}_{\boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\mathbf{A}}^{\perp}\right] = \phi_R^{-2}\left(N - (T+1)\right). \quad (113)
$$

Substituting (113) into (104) yields (74).

APPENDIX J

We define the following partitioned matrix:

$$
[\mathbf{s}|\mathbf{W}] \triangleq \left(\boldsymbol{\Phi}_{\mathbf{y}}^{-\top/2} \otimes \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right)\left[\frac{\partial\phi_y}{\partial\phi_S}\middle| \frac{\partial\phi_y}{\partial\boldsymbol{\sigma}^{\top}}, \frac{\partial\phi_y}{\partial\phi_R}\right], \quad (114)
$$

where $\boldsymbol{\sigma} \triangleq \mathrm{vec}\left(\{\phi_{SV}\} \setminus \{\phi_S\}\right)$. Similarly to (103)–(104), we use (114) to construct the FIM, and then compute its inverse and take the corresponding component,

$$
\mathrm{CRB}\left(\phi_S\right) = \frac{1}{L}\left(\mathbf{s}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{W}}^{\perp}\mathbf{s}\right)^{-1}. \quad (115)
$$

We define a partition of $\mathbf{W}$ as

$$
\mathbf{W} = \left(\boldsymbol{\Phi}_{\mathbf{y}}^{-\top/2} \otimes \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right)\left[\frac{\partial\phi_y}{\partial\boldsymbol{\sigma}^{\top}}\middle| \frac{\partial\phi_y}{\partial\phi_R}\right] \triangleq [\boldsymbol{\Sigma}|\mathbf{g}]. \quad (116)
$$

Using the blockwise formula for projection matrices [34],

$$
\boldsymbol{\Pi}_{\mathbf{W}}^{\perp} = \boldsymbol{\Pi}_{\boldsymbol{\Sigma}}^{\perp} - \boldsymbol{\Pi}_{\boldsymbol{\Sigma}}^{\perp}\mathbf{g}\left(\mathbf{g}^{\mathrm{H}}\boldsymbol{\Pi}_{\boldsymbol{\Sigma}}^{\perp}\mathbf{g}\right)^{-1}\mathbf{g}^{\mathrm{H}}\boldsymbol{\Pi}_{\boldsymbol{\Sigma}}^{\perp}. \quad (117)
$$

We note that $\boldsymbol{\Sigma}$ can be further partitioned as

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \left(\boldsymbol{\Phi}_{\mathbf{y}}^{-\top/2} \otimes \boldsymbol{\Phi}_{\mathbf{y}}^{-1/2}\right) \\
&\quad \times \left[\mathbf{V}^* \otimes \mathbf{V}\middle| \mathbf{g}_d^* \otimes \mathbf{V}, \mathbf{V}^* \otimes \mathbf{g}_d\right] \triangleq [\mathbf{C}|\mathbf{D}]. \quad (118)
\end{aligned}
$$

In a similar manner to (117), we obtain

$$
\boldsymbol{\Pi}_{\boldsymbol{\Sigma}}^{\perp} = \boldsymbol{\Pi}_{\mathbf{C}}^{\perp} - \boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{D}\left(\mathbf{D}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}. \quad (119)
$$

Substituting (119) into (117) and then into (115), yields

$$
\mathrm{CRB}\left(\phi_S\right) = \frac{1}{L}\left(\alpha_s - \boldsymbol{\alpha}_{ds}^{\mathrm{H}}\boldsymbol{\Theta}_d^{-1}\boldsymbol{\alpha}_{ds} - \beta_{sg}^2\beta_g^{-1}\right)^{-1}, \quad (120)
$$

where $\alpha_s = \mathbf{s}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{s}$, $\boldsymbol{\alpha}_{ds} = \mathbf{D}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{s}$, $\boldsymbol{\Theta}_d = \mathbf{D}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{D}$, $\beta_{sg} = \alpha_{sg} - \boldsymbol{\alpha}_{ds}^{\mathrm{H}}\boldsymbol{\Theta}_d^{-1}\boldsymbol{\alpha}_{dg}$, $\beta_g = \alpha_g - \boldsymbol{\alpha}_{dg}^{\mathrm{H}}\boldsymbol{\Theta}_d^{-1}\boldsymbol{\alpha}_{dg}$, and $\alpha_{sg} = \mathbf{s}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{g}$, $\boldsymbol{\alpha}_{dg} = \mathbf{D}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{g}$, $\alpha_g = \mathbf{g}^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{C}}^{\perp}\mathbf{g}$. These quantities can be computed using similar techniques to those used in the derivation of (106)–(113). Due to space constraints, the detailed derivation is omitted. Collecting all the terms and substituting into (120), yields the same expression as the MSE in (71).

REFERENCES

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

[2] 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.

[3] 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.

[4] 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, p. 34, 2015.

[5] O. Schwartz, S. Braun, S. Gannot, and E. A. 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)*, 2015, pp. 1–5.

[6] O. Schwartz, S. Gannot, and E. A. 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)*, 2016, pp. 151–155.

[7] A. Kuklasinski, S. Doclo, and J. Jensen, "Maximum likelihood psd estimation for speech enhancement in reverberant and noisy conditions," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 599–603.

[8] O. Schwartz, S. Gannot, and E. A. Habets, "Joint estimation of late reverberant and speech power spectral densities in noisy environments using frobenius norm," in *24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 1123–1127.

[9] 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. 1599–1612, 2016.

[10] S. Braun, A. Kuklasinski, O. Schwartz, O. Thiergart, E. A. Habets, S. Gannot, S. Doclo, and J. Jensen, "Evaluation and comparison of late reverberation power spectral density estimators," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 26, no. 6, pp. 1052–1067, 2018.

[11] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*.   Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[12] 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.

[13] 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.

[14] O. Schwartz, S. Gannot, and E. A. Habets, "Cramér–Rao bound analysis of reverberation level estimators for dereverberation and noise reduction," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 8, pp. 1680–1693, 2017.

[15] I. Kodrasi and S. Doclo, "Joint late reverberation and noise power spectal density estimation in a spatially homogenous noise field," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[16] M. Tammen, S. Doclo, and I. Kodrasi, "Joint estimation of RETF vector and power spectral densities for speech enhancement based on alternating least squares," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 795–799.

[17] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Robust joint estimation of multimicrophone signal model parameters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 7, pp. 1136–1150, 2019.

[18] R. A. Wooding, "The multivariate distribution of complex normal variables," *Biometrika*, vol. 43, no. 1/2, pp. 212–215, 1956.

[19] B. F. Cron and C. H. Sherman, "Spatial-correlation functions for various noise models," *The Journal of the Acoustical Society of America*, vol. 34, no. 11, pp. 1732–1736, 1962.

[20] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation and modulation theory*.   Wiley, 2002.

[21] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple inter-fering speech signals," *IEEE Tran. on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, 2009.

[22] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone arrays*.   Springer, 2001, pp. 39–60.

[23] R. Balan and J. Rosca, "Microphone array speech enhancement by Bayesian estimation of spectral amplitude and phase," in *IEEE Sensor Array and Multichannel Signal Process. Workshop*, 2002, pp. 209–213.

[24] 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.

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

[26] I.-T. Recommendation, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," *Rec. ITU-T P. 862*, 2001.

[27] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective measures of speech quality*.   Prentice Hall, 1988.

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

[29] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1," Disc, 1993.

[30] A. Graham, *Kronecker products and matrix calculus with applications*. Courier Dover Publications, 2018.

[31] W. J. Bangs, "Array processing with generalized beamformers," Ph.D. dissertation, Yale University, New Haven, CT, 1972.

[32] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Processing Letters*, vol. 8, no. 5, pp. 148–150, 2001.

[33] A. Gershman, P. Stoica, M. Pesavento, and E. G. Larsson, "Stochastic Cramér–Rao bound for direction estimation in unknown noise fields," *IEE Proceedings-Radar, Sonar and Navigation*, vol. 149, no. 1, pp. 2–8, 2002.

[34] C. R. Rao and H. Toutenburg, *Linear models: least squares and alternatives*.   Springer, 1999.