KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Randall Ali, Toon van Waterschoot, Marc Moonen, (2019), **Using partial a priori knowledge of relative transfer functions to design an MVDR beamformer for a binaural hearing assistive device with external microphones**
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	
Journal homepage	http://www.ica2019.org/
Author contact	your email randall.ali@esat.kuleuven.be
IR	

(article begins on next page)

PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Using partial a priori knowledge of relative transfer functions to design an MVDR beamformer for a binaural hearing assistive device with external microphones

Randall ALI[(1)], Toon vanWATERSCHOOT[(2)], Marc MOONEN[(3)]

[(1)]KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS), Leuven, Belgium, randall.ali@esat.kuleuven.be
[(2)]KU Leuven, Dept. of Electrical Engineering (ESAT-ETC/STADIUS), Leuven, Belgium, tvanwate@esat.kuleuven.be
[(3)]KU Leuven, Dept. of Electrical Engineering (ESAT-STADIUS), Leuven, Belgium, marc.moonen@esat.kuleuven.be

Abstract

This paper considers a binaural hearing assistive device (HAD) equipped with a separate local microphone array (LMA) for the left and right ear, as well as external microphones (XMs) that may be located within the vicinity of this HAD. For such a system, a binaural minimum variance distortionless response (BMVDR) beamformer may be used for noise reduction, and for the preservation of the relevant binaural speech cues, provided that a reliable estimate of the left and right ear relative transfer function (RTF) vectors pertaining to all the microphones can be obtained. In this paper, an alternative approach is considered, which makes use of available partial a priori knowledge of these RTF vectors, i.e., known separate left and right ear RTF vectors for the respective LMAs on the binaural HAD. The procedure for this approach will be discussed, which requires the estimation of an appropriate scaling between the left and right ear RTF vectors, and the missing part of these RTF vectors pertaining to the XMs. An experiment involving a dummy head, two behind-the-ear dummy hearing aids, and XMs is also performed in order to evaluate the benefit of the proposed approach.

Keywords: Binaural MVDR, External Microphones, Hearing Assistive Device

1 INTRODUCTION

In noisy environments, speech intelligibility is inevitably degraded for individuals that suffer with a hearing impairment and hence hearing assistive devices (HADs) such as hearing aids (HAs) or cochlear implants (CIs) must perform speech enhancement tasks. In addition to the fundamental task of noise reduction, preservation of the binaural cues, i.e., the interaural time differences (ITDs) and interaural level differences (ILDs) is also important to maintain the spatial perception of the auditory scene.

For a binaural HAD equipped with a separate local microphone array (LMA) for the left and right ear, and a communication link between them, the binaural minimum variance distortionless response beamformer (BMVDR) (1) is known to exhibit substantial noise reduction and to preserve the ITD and ILD of a target speaker[1]. In recent work (2, 3), such a binaural HAD has also been supplemented with an external microphone (XM) (e.g. a wearable microphone or the microphone on a mobile device) and it was demonstrated that the XM could contribute to additional noise reduction and preserve the relevant binaural cues. For the successful operation of the BMVDR in this case, an estimate of the entire vector of transfer functions from the target signal at a left ear reference microphone to all the other microphones, i.e., the left ear relative transfer function (RTF) vector, and a corresponding right ear RTF vector is required. However, obtaining such estimates becomes increasingly challenging in adverse acoustic conditions.

Therefore in this paper, generalising the system to include more than one XM, an alternative approach is considered, which makes use of available partial a priori knowledge of these RTF vectors, i.e., known separate left and right ear RTF vectors for the respective LMAs on the binaural HAD (4). In such a case, it is only the

[1] Although, the BMVDR beamformer preserves the binaural cues for the target speaker, it distorts the binaural cues for the noise. However, in (1), several remedies have been proposed, and hence this work will focus only on the preservation of the binaural cues for the target speaker.

estimation of an appropriate scaling between the left and right ear RTF vectors, and the missing part of these RTF vectors pertaining to the XMs that need to be estimated.

The paper is organised as follows. In Section 2, the data model and notation are described. In Section 3, the state of the art procedure for estimating the entire RTF vector is reviewed. In Section 4, the proposed procedure that makes use of the partial a priori knowledge of the RTF vector is discussed. In Section 5, the proposed procedure is evaluated using recorded audio data, and conclusions are drawn in Section 6.

2 DATA MODEL

The scenario as depicted in Figure 1 is considered, in which a user of a binaural HAD is listening to one target speaker of interest in a noisy, reverberant environment. The binaural HAD consists of an LMA with M_{a} microphones for the left ear and an LMA with M_{a} microphones for the right ear. Additionally, there are M_{e} XMs randomly placed within the room ($M_{\mathrm{e}} = 2$ in Fig.1). In the short-time Fourier transform (STFT) domain,

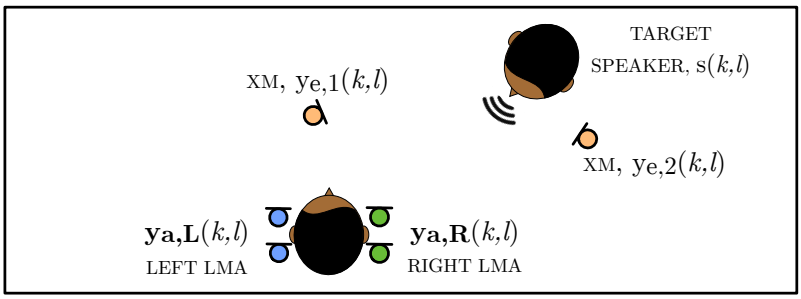

Figure 1. Scenario with a user of a binaural HAD having access to XMs, listening to the target speaker.

the microphone signals at one frequency, k, and one time frame, l, can be stacked into a vector and represented as follows:

$$\mathbf{y}(k,l) = \underbrace{\mathbf{a}(k,l)\mathrm{s}(k,l)}_{\mathbf{x}(k,l)} + \mathbf{n}(k,l) \implies \begin{bmatrix} \mathbf{y}_{\mathbf{a},\mathrm{L}}(k,l) \\ \mathbf{y}_{\mathbf{a},\mathrm{R}}(k,l) \\ \mathbf{y}_{\mathbf{e}}(k,l) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{\mathbf{a},\mathrm{L}}(k,l) \\ \mathbf{a}_{\mathbf{a},\mathrm{R}}(k,l) \\ \mathbf{a}_{\mathbf{e}}(k,l) \end{bmatrix} \mathrm{s}(k,l) + \begin{bmatrix} \mathbf{n}_{\mathbf{a},\mathrm{L}}(k,l) \\ \mathbf{n}_{\mathbf{a},\mathrm{R}}(k,l) \\ \mathbf{n}_{\mathbf{e}}(k,l) \end{bmatrix} \tag{1}$$

where[2] $\mathbf{a}_{\mathbf{a},\mathrm{L}} = [\mathrm{a}_{\mathrm{a1,L}},\ \mathrm{a}_{\mathrm{a2,L}},\ \ldots\ \mathrm{a}_{\mathrm{aM_a,L}}]^T$, $\mathbf{a}_{\mathbf{a},\mathrm{R}} = [\mathrm{a}_{\mathrm{a1,R}},\ \mathrm{a}_{\mathrm{a2,R}},\ \ldots\ \mathrm{a}_{\mathrm{aM_a,R}}]^T$, and $\mathbf{a}_{\mathbf{e}} = [\mathrm{a}_{\mathrm{e1}},\ \mathrm{a}_{\mathrm{e2}},\ \ldots\ \mathrm{a}_{\mathrm{eM_e}}]^T$ are the acoustic transfer functions (ATFs) from the target speaker to the microphones on the left LMA, the right LMA, and the XMs respectively. Furthermore, s is the target speaker, and $\mathbf{n}_{\mathbf{a},\mathrm{L}}$, $\mathbf{n}_{\mathbf{a},\mathrm{R}}$, and $\mathbf{n}_{\mathbf{e}}$ are the noise contributions similarly defined as $\mathbf{a}_{\mathbf{a},\mathrm{L}}$, $\mathbf{a}_{\mathbf{a},\mathrm{R}}$, and $\mathbf{a}_{\mathbf{e}}$ respectively. Without loss of generality, the first microphone in each of the LMAs is also chosen as the reference microphone:

$$\mathrm{y}_{\mathrm{a1,L}} = \mathbf{e}_{\mathbf{L}}^T\mathbf{y} = \mathrm{s}_{\mathrm{a1,L}} + \mathrm{n}_{\mathrm{a1,L}} \qquad \mathrm{y}_{\mathrm{a1,R}} = \mathbf{e}_{\mathbf{R}}^T\mathbf{y} = \mathrm{s}_{\mathrm{a1,R}} + \mathrm{n}_{\mathrm{a1,R}} \tag{2}$$

where $\mathrm{s}_{\mathrm{a1,L}} = \mathrm{a}_{\mathrm{a1,L}}\mathrm{s}$, $\mathrm{s}_{\mathrm{a1,R}} = \mathrm{a}_{\mathrm{a1,R}}\mathrm{s}$, which are the speech components that need to be estimated, and $\mathbf{e}_{\mathbf{L}}$ and $\mathbf{e}_{\mathbf{R}}$ are all-zero vectors except for a one in the left and right LMA reference microphone position respectively. In order to perform the estimation, it is firstly convenient to re-define eq. (1) in terms of a relative transfer function (RTF) vector, as opposed to the ATF vector, $\mathbf{a}$. The RTF vector is simply the ATF vector normalised to a reference microphone. Therefore in the binaural context, a separate RTF vector can be defined for the left ear and another for the right ear. Hence, eq. (1) can be expressed as follows:

$$\mathbf{y} = \mathbf{h}_{\mathrm{L}}\mathrm{s}_{\mathrm{a1,L}} + \mathbf{n} \qquad \mathbf{y} = \mathbf{h}_{\mathrm{R}}\mathrm{s}_{\mathrm{a1,R}} + \mathbf{n} \tag{3}$$

[2] The dependence on (k,l) is dropped for notational convenience.

where $\mathbf{h}_{\mathrm{L}}$ and $\mathbf{h}_{\mathrm{R}}$ are the RTF vectors defined as:

$$\mathbf{h}_{\mathrm{L}} = \frac{1}{\mathrm{a}_{\mathrm{a1,L}}} \begin{bmatrix} \mathbf{a}_{\mathbf{a},\mathrm{L}} \\ \mathbf{a}_{\mathbf{a},\mathrm{R}} \\ \mathbf{a}_{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{\mathbf{a},\mathrm{L}} \\ \varphi \mathbf{h}_{\mathbf{a},\mathrm{R}} \\ \mathbf{h}_{\mathbf{e},\mathrm{L}} \end{bmatrix} \qquad \mathbf{h}_{\mathrm{R}} = \frac{1}{\mathrm{a}_{\mathrm{a1,R}}} \begin{bmatrix} \mathbf{a}_{\mathbf{a},\mathrm{L}} \\ \mathbf{a}_{\mathbf{a},\mathrm{R}} \\ \mathbf{a}_{\mathbf{e}} \end{bmatrix} = \begin{bmatrix} \frac{1}{\varphi}\mathbf{h}_{\mathbf{a},\mathrm{L}} \\ \mathbf{h}_{\mathbf{a},\mathrm{R}} \\ \frac{1}{\varphi}\mathbf{h}_{\mathbf{e},\mathrm{L}} \end{bmatrix} \tag{4}$$

where $\mathbf{h}_{\mathbf{a},\mathrm{L}} = \frac{\mathbf{a}_{\mathbf{a},\mathrm{L}}}{\mathrm{a}_{\mathrm{a1,L}}}$ and $\mathbf{h}_{\mathbf{a},\mathrm{R}} = \frac{\mathbf{a}_{\mathbf{a},\mathrm{R}}}{\mathrm{a}_{\mathrm{a1,R}}}$ are the individual RTF vectors corresponding to each of the left and right LMAs and the complex scaling, $\varphi = \frac{\mathrm{a}_{\mathrm{a1,R}}}{\mathrm{a}_{\mathrm{a1,L}}}$. It should be noted that the part of the RTF vector pertaining to the XMs in $\mathbf{h}_{\mathrm{R}}$ is a scaled version of that in $\mathbf{h}_{\mathrm{L}}$, where $\mathbf{h}_{\mathbf{e},\mathrm{L}} = \frac{\mathbf{a}_{\mathbf{e}}}{\mathrm{a}_{\mathrm{a1,L}}}$. In fact, it can be seen that $\mathbf{h}_{\mathrm{L}} = \varphi \mathbf{h}_{\mathrm{R}}$, which means that the RTF vectors are parallel.

The speech-plus-noise spatial correlation matrix, $\mathbf{R}_{\mathbf{yy}}$, the noise-only correlation matrix, $\mathbf{R}_{\mathbf{nn}}$, and the speech-only correlation matrix, $\mathbf{R}_{\mathbf{xx}}$, all $\in \mathbb{C}^{(2M_{\mathrm{a}}+M_{\mathrm{e}})\times(2M_{\mathrm{a}}+M_{\mathrm{e}})}$, are given respectively as:

$$\mathbf{R}_{\mathbf{yy}} = \mathbb{E}\{\mathbf{y}\mathbf{y}^{\mathrm{H}}\}; \;\; \mathbf{R}_{\mathbf{nn}} = \mathbb{E}\{\mathbf{n}\mathbf{n}^{\mathrm{H}}\}; \;\; \mathbf{R}_{\mathbf{xx}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} \tag{5}$$

where $\mathbb{E}\{.\}$ is the expectation operator, ^{H} is the Hermitian transpose, and $\mathbf{R}_{\mathbf{xx}}$ is a rank-1 correlation matrix:

$$\mathbf{R}_{\mathbf{xx}} = \mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} = \sigma^2_{\mathrm{s_{a1,L}}} \mathbf{h}_{\mathrm{L}} \mathbf{h}_{\mathrm{L}}^{\mathrm{H}} = \sigma^2_{\mathrm{s_{a1,R}}} \mathbf{h}_{\mathrm{R}} \mathbf{h}_{\mathrm{R}}^{\mathrm{H}} \tag{6}$$

where $\sigma^2_{\mathrm{s_{a1,L}}} = \mathbb{E}\{|\mathrm{s_{a1,L}}|^2\}$ and $\sigma^2_{\mathrm{s_{a1,R}}} = \mathbb{E}\{|\mathrm{s_{a1,R}}|^2\}$ are the speech powers in the reference left and right microphone respectively. It is also assumed that the speech components are uncorrelated with the noise components, and hence $\mathbf{R}_{\mathbf{yy}} = \mathbf{R}_{\mathbf{xx}} + \mathbf{R}_{\mathbf{nn}}$. A perfect communication link is additionally assumed among the left and right LMAs in the binaural HAD, and the XMs, with no bandwidth constraints and synchronous sampling.

The estimate of the speech component in the reference microphone of the left and right LMAs, i.e. the estimate of $\mathrm{s_{a1,L}}$ and $\mathrm{s_{a1,R}}$ is then obtained through the linear filtering of the microphone signals, with the complex-valued filters, $\mathbf{w}_{\mathrm{L}}$ and $\mathbf{w}_{\mathrm{R}}$ respectively:

$$\hat{\mathrm{s}}_{\mathrm{a1,L}} = \mathbf{w}_{\mathrm{L}}^{\mathrm{H}} \mathbf{y} \qquad \hat{\mathrm{s}}_{\mathrm{a1,R}} = \mathbf{w}_{\mathrm{R}}^{\mathrm{H}} \mathbf{y} \tag{7}$$

The BMVDR beamformer filters, $\mathbf{w}_{\mathrm{L}}$ and $\mathbf{w}_{\mathrm{R}}$, are then given by:

$$\mathbf{w}_{\mathrm{L}} = \frac{\mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{h}_{\mathrm{L}}}{\mathbf{h}_{\mathrm{L}}^{\mathrm{H}} \mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{h}_{\mathrm{L}}} \qquad \mathbf{w}_{\mathrm{R}} = \frac{\mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{h}_{\mathrm{R}}}{\mathbf{h}_{\mathrm{R}}^{\mathrm{H}} \mathbf{R}_{\mathbf{nn}}^{-1} \mathbf{h}_{\mathrm{R}}} \tag{8}$$

Consequently, in order to compute these filters, estimates are required for $\mathbf{R}_{\mathbf{nn}}$, and the RTFs, $\mathbf{h}_{\mathrm{L}}$, and $\mathbf{h}_{\mathrm{R}}$. Typically, $\mathbf{R}_{\mathbf{nn}}$ can be estimated during periods of noise only with recursive averaging (3). Hence this paper focuses on the estimation of $\mathbf{h}_{\mathrm{L}}$ and $\mathbf{h}_{\mathrm{R}}$.

3 ESTIMATING THE ENTIRE RTF VECTOR

Given $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$, which are estimates of $\mathbf{R}_{\mathbf{yy}}$ and $\mathbf{R}_{\mathbf{nn}}$ respectively, a generalised eigenvalue decomposition (GEVD) (5) or what is equivalently known as covariance whitening (3) can be used to estimate $\mathbf{h}_{\mathrm{L}}$ and $\mathbf{h}_{\mathrm{R}}$. A spatial pre-whitening operation can be firstly defined from $\hat{\mathbf{R}}_{\mathbf{nn}}$ using the Cholesky decomposition:

$$\hat{\mathbf{R}}_{\mathbf{nn}} = \hat{\mathbf{R}}_{\mathbf{nn}}^{1/2} \, \hat{\mathbf{R}}_{\mathbf{nn}}^{\mathrm{H}/2} \tag{9}$$

where $\hat{\mathbf{R}}_{\mathbf{nn}}^{1/2}$ is a lower triangular matrix. Spatial pre-whitening is then performed by pre-multiplying the signal vector of interest by $\hat{\mathbf{R}}_{\mathbf{nn}}^{-1/2}$. For an autocorrelation matrix, spatial pre-whitening is performed by pre-multiplying it by $\hat{\mathbf{R}}_{\mathbf{nn}}^{-1/2}$ and post-multiplying it by $\hat{\mathbf{R}}_{\mathbf{nn}}^{-\mathrm{H}/2}$. Using the definition of $\mathbf{R}_{\mathbf{xx}}$ from eq. (6) and that $\mathbf{R}_{\mathbf{yy}} = \mathbf{R}_{\mathbf{xx}} + \mathbf{R}_{\mathbf{nn}}$, the following optimisation problem can be considered to estimate $\mathbf{h}_{\mathrm{L}}$ (and $\mathbf{h}_{\mathrm{R}}$ by an appropriate scaling):

$$\min_{\sigma^2_{\mathrm{s_{a1,L}}}, \mathbf{h}_{\mathrm{L}}} ||\hat{\mathbf{R}}_{\mathbf{nn}}^{-1/2}((\hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{nn}}) - \sigma^2_{\mathrm{s_{a1,L}}} \mathbf{h}_{\mathrm{L}} \mathbf{h}_{\mathrm{L}}^{\mathrm{H}}) \hat{\mathbf{R}}_{\mathbf{nn}}^{-\mathrm{H}/2}||_F^2 \tag{10}$$

where $||.||_F$ is the frobenius norm. The solution to eq. (10) then follows from an eigenvalue decomposition (EVD) of $\hat{\mathbf{R}}_{\mathbf{nn}}^{-\mathrm{H}/2}\hat{\mathbf{R}}_{\mathbf{yy}}\hat{\mathbf{R}}_{\mathbf{nn}}^{-\mathrm{H}/2}$ or equivalently, GEVD of the matrix pencil $\{\hat{\mathbf{R}}_{\mathbf{yy}}, \hat{\mathbf{R}}_{\mathbf{nn}}\}$:

$$\hat{\mathbf{R}}_{\mathbf{nn}}^{-1}\hat{\mathbf{R}}_{\mathbf{yy}} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^{-1} \tag{11}$$

where $\mathbf{\Sigma}$ is a diagonal matrix of the generalised eigenvalues arranged in descending order, and $\mathbf{U}$ is an invertible matrix containing the corresponding generalised eigenvectors. The GEVD is also equivalent to a joint diagonalisation of $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$:

$$\hat{\mathbf{R}}_{\mathbf{yy}} = \mathbf{Q}\mathbf{\Sigma}_{\mathbf{y}}\mathbf{Q}^{\mathrm{H}} \qquad \hat{\mathbf{R}}_{\mathbf{nn}} = \mathbf{Q}\mathbf{\Sigma}_{\mathbf{n}}\mathbf{Q}^{\mathrm{H}} \tag{12}$$

where $\mathbf{\Sigma}_{\mathbf{y}}$ and $\mathbf{\Sigma}_{\mathbf{n}}$ are diagonal matrices, and $\mathbf{Q} = \mathbf{U}^{-\mathrm{H}}$ is an invertible matrix. A rank-1 approximation to $(\hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{nn}}) = \mathbf{Q}(\mathbf{\Sigma}_{\mathbf{y}} - \mathbf{\Sigma}_{\mathbf{n}})\mathbf{Q}^{\mathrm{H}}$ yields an estimate for $\mathbf{R}_{\mathbf{xx}}$, $\hat{\mathbf{R}}_{\mathbf{xx}} = \mathbf{Q}\mathbf{e_1}\,\mathbf{e_1}^T(\mathbf{\Sigma}_{\mathbf{y}} - \mathbf{\Sigma}_{\mathbf{n}})\mathbf{e_1}\,\mathbf{e_1}^T\mathbf{Q}^{\mathrm{H}}$, where $\mathbf{e_1} \in \mathbb{C}^{2M_\mathrm{a}+M_\mathrm{e}}$ is an all-zero vector except for a one as the first element (and it is noted that $\mathbf{e_1} = \mathbf{e_L}$). It can be shown (5) that this corresponds to the rank-1 approximation sought from eq. (10) so that the estimates to $\mathbf{h}_\mathrm{L}$ and $\mathbf{h}_\mathrm{R}$ then follow as:

$$\hat{\mathbf{h}}_\mathrm{L} = \frac{\mathbf{Q}\mathbf{e_1}}{\mathbf{e}_\mathbf{L}^T\mathbf{Q}\mathbf{e_1}} \qquad \hat{\mathbf{h}}_\mathrm{R} = \frac{\mathbf{Q}\mathbf{e_1}}{\mathbf{e}_\mathbf{R}^T\mathbf{Q}\mathbf{e_1}} \tag{13}$$

Finally, a substitution of $\hat{\mathbf{R}}_{\mathbf{nn}}$ from eq. (12) and $\hat{\mathbf{h}}_\mathrm{L}$ and $\hat{\mathbf{h}}_\mathrm{R}$ from eq. (13) into eq. (8) results in the corresponding BMVDR filters:

$$\hat{\mathbf{w}}_\mathrm{L} = \mathbf{U}\mathbf{e_1}\,\mathbf{e_1}^T\mathbf{Q}^{\mathrm{H}}\mathbf{e_L} \qquad \hat{\mathbf{w}}_\mathrm{R} = \mathbf{U}\mathbf{e_1}\,\mathbf{e_1}^T\mathbf{Q}^{\mathrm{H}}\mathbf{e_R} \tag{14}$$

4 USING PARTIAL A PRIORI KNOWLEDGE OF THE RTF VECTOR

As opposed to estimating the entire RTF vectors, $\mathbf{h}_\mathrm{L}$, and $\mathbf{h}_\mathrm{R}$, an alternative procedure may be followed if there is a priori knowledge of the RTF vectors for the separate left and right LMA, i.e., if a suitable approximation to $\mathbf{h}_{\mathbf{a},\mathrm{L}}$ and $\mathbf{h}_{\mathbf{a},\mathrm{R}}$ is available. For instance, such an approximation may be the measured RTF vectors for the separate left and right LMA in an anechoic room or RTF vectors from an existing binaural noise reduction system that uses only the LMAs. Denoting this approximation to $\mathbf{h}_{\mathbf{a},\mathrm{L}}$ and $\mathbf{h}_{\mathbf{a},\mathrm{R}}$ as $\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{L}}$ and $\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{R}}$ respectively, and recalling the definitions from eq. (4), an alternative optimisation problem to eq. (10) can be considered:

$$\min_{\sigma^2_{\mathrm{s_{a1,L}}},\,\varphi,\,\mathbf{h}_{\mathbf{e},\mathrm{L}}} ||\hat{\mathbf{R}}_{\mathbf{nn}}^{-\mathrm{H}/2}((\hat{\mathbf{R}}_{\mathbf{yy}} - \hat{\mathbf{R}}_{\mathbf{nn}}) - \sigma^2_{\mathrm{s_{a1,L}}}\begin{bmatrix}\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{L}}\\ \varphi\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{R}}\\ \mathbf{h}_{\mathbf{e},\mathrm{L}}\end{bmatrix}\begin{bmatrix}\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{L}}^{\mathrm{H}} & \varphi^*\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{R}}^{\mathrm{H}} & \mathbf{h}_{\mathbf{e},\mathrm{L}}^{\mathrm{H}}\end{bmatrix})\hat{\mathbf{R}}_{\mathbf{nn}}^{-\mathrm{H}/2}||_F^2 \tag{15}$$

where now it is only the scaling, φ, and the RTF vector for the XMs, $\mathbf{h}_{\mathbf{e},\mathrm{L}}$, which need to be found as opposed to the entire $\mathbf{h}_\mathrm{L}$ as in eq. (10). As will be discussed in the following, the solution can be realised in the block scheme of Figure 2, which consists of compressing the left and right LMA signals, an orthogonalisation operation, and finally a GEVD on a lower dimensional ($\mathbb{C}^{(M_\mathrm{e}+2)\times(M_\mathrm{e}+2)}$) matrix pencil. In order to solve eq. (15), the following blocking matrices, $\mathbf{C_a} \in \mathbb{C}^{2M_\mathrm{a}\times(2M_\mathrm{a}-2)}$, $\mathbf{C}_{\mathbf{a},\mathrm{L}} \in \mathbb{C}^{M_\mathrm{a}\times(M_\mathrm{a}-1)}$, $\mathbf{C}_{\mathbf{a},\mathrm{R}} \in \mathbb{C}^{M_\mathrm{a}\times(M_\mathrm{a}-1)}$, fixed beamformers, $\mathbf{F_a} \in \mathbb{C}^{2M_\mathrm{a}\times 2}$, $\mathbf{f}_{\mathbf{a},\mathrm{L}} \in \mathbb{C}^{M_\mathrm{a}}$, $\mathbf{f}_{\mathbf{a},\mathrm{R}} \in \mathbb{C}^{M_\mathrm{a}}$, and transformation matrix, $\mathbf{T} \in \mathbb{C}^{(2M_\mathrm{a}+M_\mathrm{e})\times(2M_\mathrm{a}+M_\mathrm{e})}$ are firstly defined:

$$\mathbf{C_a} = \begin{bmatrix}\mathbf{C}_{\mathbf{a},\mathrm{L}} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}_{\mathbf{a},\mathrm{R}}\end{bmatrix} \qquad \mathbf{F_a} = \begin{bmatrix}\mathbf{f}_{\mathbf{a},\mathrm{L}} & \mathbf{0}\\ \mathbf{0} & \mathbf{f}_{\mathbf{a},\mathrm{R}}\end{bmatrix} \qquad \mathbf{T} = \left[\begin{array}{c|c|c}\mathbf{C_a} & \mathbf{F_a} & \mathbf{0}\\ \hline \mathbf{0} & \mathbf{0} & \mathbf{I}_{M_\mathrm{e}}\end{array}\right]$$
$$\mathbf{C}_{\mathbf{a},\mathrm{L}}^{\mathrm{H}}\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{L}} = \mathbf{0};\ \mathbf{C}_{\mathbf{a},\mathrm{R}}^{\mathrm{H}}\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{R}} = \mathbf{0} \qquad \mathbf{f}_{\mathbf{a},\mathrm{L}}^{\mathrm{H}}\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{L}} = 1;\ \mathbf{f}_{\mathbf{a},\mathrm{R}}^{\mathrm{H}}\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{R}} = 1 \tag{16}$$

where $\mathbf{I}_{M_\mathrm{e}} \in \mathbb{C}^{M_\mathrm{e}\times M_\mathrm{e}}$ is an identity matrix. The first two blocks in Fig. 2 apply the transformation, $\mathbf{T}^{\mathrm{H}}$, to $\mathbf{y}$ to yield a set of blocking matrix signals, $\mathbf{C}_{\mathbf{a}}^{\mathrm{H}}\mathbf{y_a} \in \mathbb{C}^{2M_\mathrm{a}-2}$, two compressed signals, $\mathbf{f}_{\mathbf{a},\mathrm{L}}^{\mathrm{H}}\mathbf{y}_{\mathbf{a},\mathrm{L}}$ and $\mathbf{f}_{\mathbf{a},\mathrm{R}}^{\mathrm{H}}\mathbf{y}_{\mathbf{a},\mathrm{R}}$ resulting from the left and right fixed beamformers respectively, and the unaltered set of XM signals, $\mathbf{y_e}$. An alternative spatial pre-whitening operation can then be defined by applying the transformation to $\hat{\mathbf{R}}_{\mathbf{nn}}$:

$$\mathbf{T}^{\mathrm{H}}\hat{\mathbf{R}}_{\mathbf{nn}}\mathbf{T}^{\mathrm{H}} = \mathbf{L}\mathbf{L}^{\mathrm{H}} \tag{17}$$

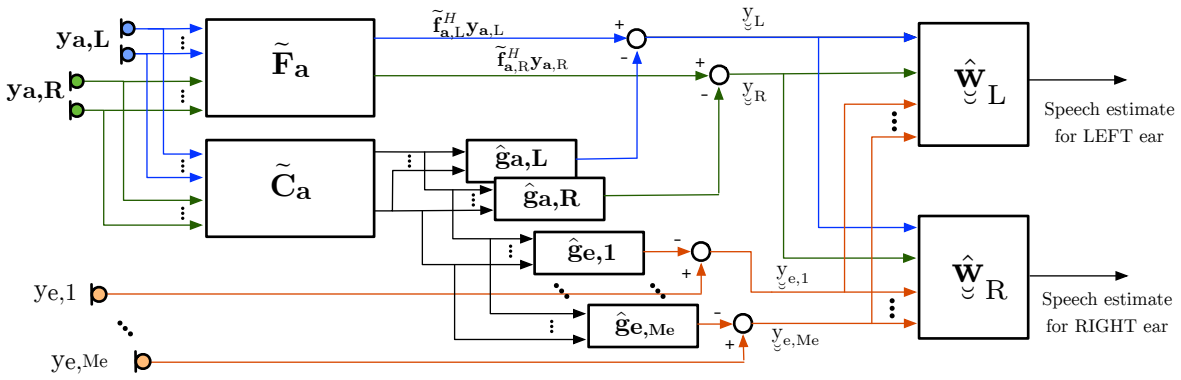

Figure 2. Block scheme for a BMVDR that uses partial a priori knowledge of the RTF vectors.

where $\mathbf{L}$ is a lower triangular matrix. It can then be shown (6) that eq. (15) can be equivalently re-written as:

$$\min_{\sigma^2_{s_{a1,L}},\,\varphi,\,\mathbf{h}_{e,L}} ||\mathbf{L}^{-1}\mathbf{T}^{H}((\hat{\mathbf{R}}_{yy}-\hat{\mathbf{R}}_{nn})-\sigma^2_{s_{a1,L}}\begin{bmatrix}\widetilde{\mathbf{h}}_{a,L}\\ \varphi\widetilde{\mathbf{h}}_{a,R}\\ \mathbf{h}_{e,L}\end{bmatrix}\begin{bmatrix}\widetilde{\mathbf{h}}^{H}_{a,L} & \varphi^{*}\widetilde{\mathbf{h}}^{H}_{a,R} & \mathbf{h}^{H}_{e,L}\end{bmatrix})\mathbf{T}\,\mathbf{L}^{-H}||^2_F \tag{18}$$

from which the solution follows eventually from a GEVD of a matrix pencil consisting of the lower dimensional $\mathbb{C}^{(M_e+2)\times(M_e+2)}$ correlation matrices (6) (7) corresponding to the M_e+2 compressed signals, $\underline{\underline{\mathbf{y}}} = [\underline{\underline{y}}_L\ \underline{\underline{y}}_R\ \underline{\underline{y}}_{e,1}\cdots\underline{\underline{y}}_{e,Me}]^T \in \mathbb{C}^{M_e+2}$. These compressed signals can be computed by performing an orthogonalisation involving the previously transformed signals, or equivalently, with a GSC beamformer, $\mathbf{A}$:

$$\underline{\underline{\mathbf{y}}} = \mathbf{A}^{H}\mathbf{y} = \mathbf{F}^{H}\mathbf{y} - \mathbf{G}^{H}\mathbf{C}_a^{H}\mathbf{E}^{T}\mathbf{y} \tag{19}$$

with $\mathbf{A}=\mathbf{F}-\mathbf{E}\mathbf{C}_a\mathbf{G}$, $\mathbf{F}\in\mathbb{C}^{(2M_a+M_e)\times(M_e+2)}$ defined as $\mathbf{F} = \begin{bmatrix}\mathbf{F}_a & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{M_e}\end{bmatrix}$, the selection matrix, $\mathbf{E}=[\mathbf{I}_{2M_a}\,|\,\mathbf{0}_{(2M_a\times M_e)}]^T$, and $\mathbf{G}\in\mathbb{C}^{(2M_a-2)\times(M_e+2)}$, which can be computed as follows:

$$\mathbf{G} = (\mathbf{C}_a^{H}\mathbf{E}^{T}\hat{\mathbf{R}}_{nn}\mathbf{E}\mathbf{C}_a)^{-1}\mathbf{C}_a^{H}\mathbf{E}^{T}\hat{\mathbf{R}}_{nn}\mathbf{F} = [\mathbf{g}_{a,L}\ \mathbf{g}_{a,R}\ \mathbf{g}_{e,1}\ \cdots\ \mathbf{g}_{e,M_e}] \tag{20}$$

is comprised of the $\mathbb{C}^{M_e+2}$ orthogonalisation filters in each column. As depicted in Fig. 2, $\mathbf{g}_{a,L}$ and $\mathbf{g}_{a,R}$ will orthogonalise the noise components of the signals $\mathbf{f}^{H}_{a,L}\mathbf{y}_{a,L}$ and $\mathbf{f}^{H}_{a,R}\mathbf{y}_{a,R}$ onto the noise components of the blocking matrix signals, and $[\mathbf{g}_{e,1}\ \cdots\ \mathbf{g}_{e,M_e}]$ will orthogonalise the noise components of each of the respective XMs onto the noise components of the blocking matrix signals.

A new signal model that resembles eq. (3) can in fact then be realised by substituting eq. (3) into eq. (19):

$$\underline{\underline{\mathbf{y}}} = \underline{\underline{\mathbf{h}}}_L s_{a1,L} + \underline{\underline{\mathbf{n}}} \qquad \underline{\underline{\mathbf{y}}} = \underline{\underline{\mathbf{h}}}_R s_{a1,R} + \underline{\underline{\mathbf{n}}} \tag{21}$$

where $\underline{\underline{\mathbf{h}}}_L = \mathbf{A}^{H}\mathbf{h}_L = \left[1\ \varphi\ \mathbf{h}^{T}_{e,L}\right]^T$, $\underline{\underline{\mathbf{h}}}_R = \mathbf{A}^{H}\mathbf{h}_R = \left[\frac{1}{\varphi}\ 1\ \frac{1}{\varphi}\mathbf{h}^{T}_{e,L}\right]^T$, $\underline{\underline{\mathbf{n}}} = \mathbf{A}^{H}\mathbf{n}$, and the associated correlation matrices, $\underline{\underline{\mathbf{R}}}_{yy} = \mathbb{E}\{\underline{\underline{\mathbf{y}}}\underline{\underline{\mathbf{y}}}^{H}\}$ and $\underline{\underline{\mathbf{R}}}_{nn} = \mathbb{E}\{\underline{\underline{\mathbf{n}}}\underline{\underline{\mathbf{n}}}^{H}\}$. These compressed signals, $\underline{\underline{\mathbf{y}}}$, can now be used to design the BMVDR as opposed to $\mathbf{y}$. Estimates of $s_{a1,L}$ and $s_{a1,R}$ can be obtained by directly filtering $\underline{\underline{\mathbf{y}}}$ with the BMVDR filters, $\underline{\underline{\mathbf{w}}}_L \in \mathbb{C}^{M_e+2}$ and $\underline{\underline{\mathbf{w}}}_R \in \mathbb{C}^{M_e+2}$ respectively (as in eq. (7) and depicted in Fig. 2):

$$\underline{\underline{\mathbf{w}}}_L = \frac{\underline{\underline{\mathbf{R}}}_{nn}^{-1}\underline{\underline{\mathbf{h}}}_L}{\underline{\underline{\mathbf{h}}}_L^{H}\underline{\underline{\mathbf{R}}}_{nn}^{-1}\underline{\underline{\mathbf{h}}}_L} \qquad \underline{\underline{\mathbf{w}}}_R = \frac{\underline{\underline{\mathbf{R}}}_{nn}^{-1}\underline{\underline{\mathbf{h}}}_R}{\underline{\underline{\mathbf{h}}}_R^{H}\underline{\underline{\mathbf{R}}}_{nn}^{-1}\underline{\underline{\mathbf{h}}}_R} \tag{22}$$

Hence, similar to eq. (8), estimates are now required for $\underline{\underline{\mathbf{R}}}_{\mathbf{nn}}$, $\underline{\mathbf{h}}_{\mathrm{L}}$, and $\underline{\mathbf{h}}_{\mathrm{R}}$ in order to compute these filters. With the estimates of the correlation matrices, $\hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{yy}} = \mathbf{A}^{\mathrm{H}}\hat{\mathbf{R}}_{\mathbf{yy}}\mathbf{A}$ and $\hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{nn}} = \mathbf{A}^{\mathrm{H}}\hat{\mathbf{R}}_{\mathbf{nn}}\mathbf{A}$, the GEVD procedure from Section 3 then follows from the matrix pencil $\{\hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{yy}}, \hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{nn}}\}$ to estimate $\underline{\mathbf{h}}_{\mathrm{L}}$ and $\underline{\mathbf{h}}_{\mathrm{R}}$:

$$\hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{nn}}^{-1}\hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{yy}} = \underline{\underline{\mathbf{U}}}\,\underline{\underline{\mathbf{\Sigma}}}\,\underline{\underline{\mathbf{U}}}^{-1}; \qquad \hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{yy}} = \underline{\underline{\mathbf{Q}}}\,\underline{\underline{\mathbf{\Sigma}}}_{\mathbf{y}}\,\underline{\underline{\mathbf{Q}}}^{\mathrm{H}}; \qquad \hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{nn}} = \underline{\underline{\mathbf{Q}}}\,\underline{\underline{\mathbf{\Sigma}}}_{\mathbf{n}}\,\underline{\underline{\mathbf{Q}}}^{\mathrm{H}} \tag{23}$$

where $\underline{\underline{\mathbf{\Sigma}}}$ is a diagonal matrix of the generalised eigenvalues arranged in descending order, and $\underline{\underline{\mathbf{U}}}$ contains the corresponding generalised eigenvectors, $\underline{\underline{\mathbf{\Sigma}}}_{\mathbf{y}}$, and $\underline{\underline{\mathbf{\Sigma}}}_{\mathbf{n}}$ are also diagonal matrices, and $\underline{\underline{\mathbf{Q}}} = \underline{\underline{\mathbf{U}}}^{-\mathrm{H}}$ is an invertible matrix. The estimates for $\underline{\mathbf{h}}_{\mathrm{L}}$ and $\underline{\mathbf{h}}_{\mathrm{R}}$ then follow as:

$$\hat{\underline{\mathbf{h}}}_{\mathrm{L}} = \frac{\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}_{\mathbf{1}}}{\underline{\mathbf{e}}_{\mathbf{L}}^{T}\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}_{\mathbf{1}}} \qquad \hat{\underline{\mathbf{h}}}_{\mathrm{R}} = \frac{\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}_{\mathbf{1}}}{\underline{\mathbf{e}}_{\mathbf{R}}^{T}\underline{\underline{\mathbf{Q}}}\,\underline{\mathbf{e}}_{\mathbf{1}}} \tag{24}$$

where the selection vectors, $\underline{\mathbf{e}}_{\mathbf{1}} = \underline{\mathbf{e}}_{\mathbf{L}} = [1,\ 0,\ \mathbf{0}_{(1\times M_{\mathrm{e}})}]^{T}$ and $\underline{\mathbf{e}}_{\mathbf{R}} = [0,\ 1,\ \mathbf{0}_{(1\times M_{\mathrm{e}})}]^{T}$. Finally, the substitution of $\hat{\underline{\underline{\mathbf{R}}}}_{\mathbf{nn}}$ from eq. (23) and $\hat{\underline{\mathbf{h}}}_{\mathrm{L}}$ and $\hat{\underline{\mathbf{h}}}_{\mathrm{R}}$ from eq. (24) into eq. (22) results in the corresponding BMVDR filters:

$$\hat{\underline{\mathbf{w}}}_{\mathrm{L}} = \underline{\underline{\mathbf{U}}}\,\underline{\mathbf{e}}_{\mathbf{1}}\,\underline{\mathbf{e}}_{\mathbf{1}}^{T}\underline{\underline{\mathbf{Q}}}^{\mathrm{H}}\underline{\mathbf{e}}_{\mathbf{L}} \qquad \hat{\underline{\mathbf{w}}}_{\mathrm{R}} = \underline{\underline{\mathbf{U}}}\,\underline{\mathbf{e}}_{\mathbf{1}}\,\underline{\mathbf{e}}_{\mathbf{1}}^{T}\underline{\underline{\mathbf{Q}}}^{\mathrm{H}}\underline{\mathbf{e}}_{\mathbf{R}} \tag{25}$$

5 RESULTS AND DISCUSSION

In order to evaluate these algorithms, audio recordings were made in an L-shaped room of height 3.8 m, whose longer dimensions were approximately 6 m × 5.5 m, with an estimated broadband reverberation time of 1.5 s. Similar to Fig. 1, a Neumann KU-100 dummy head was placed in a central location of the room and equipped with two behind-the-ear (BTE) hearing aids (HAs), each consisting of an LMA with two microphones spaced approximately 1.3 cm apart. Denoting 0° as the azimuth direction directly in front of the dummy head and positive angles as clockwise, two XMs (AKG-CK-97-O) were positioned such that one was at 60° and 1 m away from the dummy head, and the other was at 20 cm below the dummy head. A Genelec 8030C loudspeaker, 1 m from the dummy head, was used to generate the speech signal from a target male speaker (8). Four of the same loudspeakers, facing away from the dummy head toward opposite corners, were used to generate uncorrelated excerpts of babble noise to approximate a diffuse noise field. The speech and noises were recorded separately, but mixed together at an input signal-to-noise ratio (SNR) of 0 dB at the reference (frontal) microphone on the right LMA.

Four algorithms were then evaluated: (i) the BMVDR from Section 3 (BMVDR$-\hat{\mathbf{h}}$), (ii) the BMVDR from Section 4 (BMVDR$-\widetilde{\mathbf{h}}$), (iii) the BMVDR from Section 3 but only using the left and right LMA (BMVDR$-\hat{\mathbf{h}}_{\mathbf{a}}$), and (iv) the BMVDR as from Section 4 but only using the left and right LMA (BMVDR$-\widetilde{\mathbf{h}}_{\mathbf{a}}$). For the processing of the algorithms, the Weighted Overlap and Add (WOLA) method (9), with a Discrete Fourier Transform (DFT) size of 128 samples, 50 % overlap, and a sampling frequency of 16 kHz was used. Using the speech presence probability (SPP) (10), periods for which the speech was active were extracted if the SPP > 0.6, and not active if the SPP < 0.4. $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$ were then estimated accordingly via recursive averaging with an averaging time of 1 s. Finally, the a priori RTFs, $\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{L}}$ and $\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{R}}$ were computed from the DFT of 128 samples (including only the direct path component) of the measured impulse responses (IRs) from the target speaker position to each of the left and right LMAs. The metrics used to evaluate the following experiments were the change in speech intelligibility-weighted SNR improvement (11) (ΔSI-SNR) from the input SI-SNR at LM1, and change in binaural cue errors, ΔITD and ΔILD, using the auditory model from (12). As the room was quite reverberant, the reference binaural signal for computing ΔITD and ΔILD was the speech signal of the target speaker convolved with the time domain direct path IRs used to define $\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{L}}$ and $\widetilde{\mathbf{h}}_{\mathbf{a},\mathrm{R}}$. All metrics were computed in 2 s time frames with a 50 % overlap.

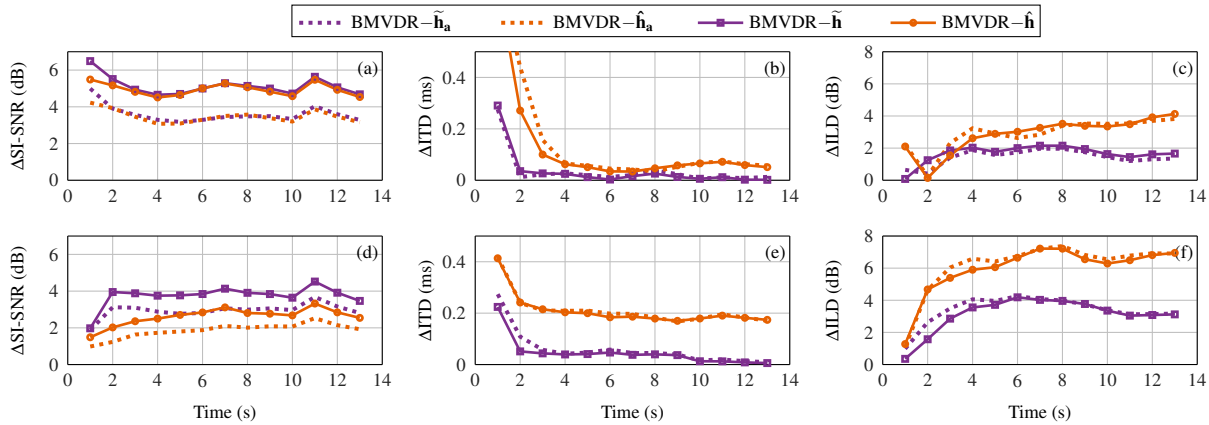

Figure 3. Performance of the various BMVDR algorithms for a target speaker located at 45°. The upper plots ((a),(b),(c)) are the results for accurately estimated correlation matrices, while the lower plots ((d),(e),(f)) are the results for less accurately estimated correlation matrices.

Figure 3 displays the results for an input signal of 13 s when the target speaker was at 45°. The upper plots ((a),(b),(c)) display the results when the SPP was computed on the speech-only signal as received by the reference microphone of the right LMA, and represents a case when $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$ have been accurately estimated. In the lower plots ((d),(e),(f)), the SPP was computed using the actual noisy signal as received by the same microphone and represents a case when $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$ have not been as accurately estimated.

Firstly, in terms of ΔSI-SNR, it can be observed that the BMVDR−$\tilde{\mathbf{h}}$ and BMVDR−$\hat{\mathbf{h}}$ (i.e., algorithms that use the LMAs along with the XMs) always demonstrate an improvement over BMVDR−$\tilde{\mathbf{h}}_\mathbf{a}$ and BMVDR−$\hat{\mathbf{h}}_\mathbf{a}$ (i.e., algorithms that use the LMAs only). Furthermore, the ΔITD and ΔILD remain relatively constant between BMVDR−$\tilde{\mathbf{h}}_\mathbf{a}$ and BMVDR−$\tilde{\mathbf{h}}$, as well as between BMVDR−$\hat{\mathbf{h}}$ and BMVDR−$\hat{\mathbf{h}}$, suggesting that the inclusion of the XMs does not contribute to any additional distortion in the binaural cues.

The impact of accurate estimation of $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$ is also evident. In the upper plots ((a),(b),(c)), where $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$ have been accurately estimated, both the BMVDR−$\tilde{\mathbf{h}}$ and BMVDR−$\hat{\mathbf{h}}$ perform quite similarly. In the lower plots ((d),(e),(f)), where $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$ have been estimated less accurately, the absolute values of all metrics now indicate a generally lower performance for all algorithms. However, in such a case, it is seen that the BMVDR−$\tilde{\mathbf{h}}$ offers an improved ΔSI-SNR and maintenance of the binaural cues in comparison with the BMVDR−$\hat{\mathbf{h}}$. Therefore, in cases of inaccurate estimation of $\hat{\mathbf{R}}_{\mathbf{yy}}$ and $\hat{\mathbf{R}}_{\mathbf{nn}}$, the approach proposed in this paper is seen to be beneficial. The resulting audio files from this experiment may be listened to for subjective evaluation (13).

6 CONCLUSIONS

An approach to designing a BMVDR beamformer for a binaural HAD consisting of separate LMAs for the left and right ear, as well as XMs has been developed. As opposed to the conventional approach of estimating entire RTF vectors, partial a priori knowledge of these RTF vectors, i.e., known separate left and right ear RTF vectors for the LMAs on the binaural HAD was incorporated so that only the computation of an appropriate scaling and the missing part of these RTF vectors pertaining to the XMs was required. An experiment with a dummy head, two behind-the-ear dummy hearing aids, and XMs indicated that the proposed approach is especially beneficial in cases where there may be inaccurate estimation of the relevant correlation matrices needed for the BMVDR beamformer.

ACKNOWLEDGEMENTS

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven C2-16-00449 'Distributed Digital Signal Processing for Ad-hoc Wireless Local Area Audio Networking', and KU Leuven Internal Funds VES/19/004. The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information. The scientific responsibility is assumed by its authors.

REFERENCES

[1] Doclo, S.; Gannot, S.; Marquardt, D.; Hadad, E. Ch. 18: Binaural Speech Processing with Application to Hearing Devices, Audio Source Separation and Speech Enhancement, Wiley, 1st Edition, 2018.

[2] Szurley J.; Bertrand A.; van Dijk B.; Moonen M. Binaural noise cue preservation in a binaural noise reduction system with a remote microphone signal. IEEE/ACM Trans. Audio Speech Lang. Process., 24(5), 2016, pp 952–966.

[3] Gößling, N.; Doclo, S. RTF-Based Binaural MVDR Beamformer Exploiting an External Microphone in a Diffuse Noise Field. Proc. ITG Conference on Speech Communication, Oldenburg, Germany, 2018.

[4] Koutrouvelis A.; Hendriks R.; Heusdens R.; Jensen J,; Guo M. Binaural beamforming using pre-determined relative acoustic transfer functions. Proc. 25th European Signal Process. Conf. (EUSIPCO '17), Kos island, Greece, 2017.

[5] Serizel R.; Moonen M.; van Dijk B,; Wouters J. Low-rank Approximation Based Multichannel Wiener Filter Algorithms for Noise Reduction with Application in Cochlear Implants. IEEE/ACM Trans Audio, Speech, Lang. Process., 22(4), 2014, pp 785–99.

[6] Ali R.; Bernardi G.; van Waterschoot T.; Moonen M. Methods of extending a generalised sidelobe canceller with external microphones. IEEE/ACM Trans Audio, Speech, Lang. Process, to appear, 2019.

[7] Van Rompaey, R.; Moonen, M. Distributed Adaptive Node-Specific Signal Estimation in Wireless Sensor Networks with Partial Prior Knowledge of the Desired Source Steering Vector. Internal Report 19-29, ESAT-STADIUS, KU Leuven (Leuven, Belgium). `ftp://ftp.esat.kuleuven.be/sista/rvanromp/19-29.pdf`.

[8] Bang and Olufsen; Music for Archimedes; CD B&O 101

[9] Crochiere, R. A weighted overlap-add method of short-time Fourier analysis/synthesis. IEEE Trans Acoust., Speech, Signal Process., 28(1), 1980, pp 99–102.

[10] Gerkmann T.; Hendriks R.; Noise power estimation based on the probability of speech presence. Proc. 2011 IEEE Workshop Appls. Signal Process. Audio Acoust. (WASPAA '11), New Paltz, NY, 2011.

[11] Greenberg J; Peterson P.; Zurek P. Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. J. Acoust. Soc. Amer., 94(5), 1993, pp 3009–3010.

[12] Dietz M.; Ewert S.; Hohmann V. Auditory model based direction estimation of concurrent speakers from binaural signals. Speech Communication, 53(5), 2011, pp 592–605.

[13] Ali R. `ftp://ftp.esat.kuleuven.be/pub/SISTA/rali/Reports/ICA2019/AudioBIN`, 2019.