The Binaural LCMV Beamformer and its Performance Analysis

Elior Hadad, *Student Member, IEEE*, Simon Doclo, *Senior Member, IEEE*, and Sharon Gannot, *Senior Member, IEEE*

***Abstract*—The recently proposed binaural linearly constrained minimum variance (BLCMV) beamformer is an extension of the well-known binaural minimum variance distortionless response (MVDR) beamformer, imposing constraints for both the desired and the interfering sources. Besides its capabilities to reduce interference and noise, it also enables to preserve the binaural cues of both the desired and interfering sources, hence making it particularly suitable for binaural hearing aid applications. In this paper, a theoretical analysis of the BLCMV beamformer is presented. In order to gain insights into the performance of the BLCMV beamformer, several decompositions are introduced that reveal its capabilities in terms of interference and noise reduction, while controlling the binaural cues of the desired and the interfering sources. When setting the parameters of the BLCMV beamformer, various considerations need to be taken into account, e.g. based on the amount of interference and noise reduction and the presence of estimation errors of the required relative transfer functions (RTFs). Analytical expressions for the performance of the BLCMV beamformer in terms of noise reduction, interference reduction, and cue preservation are derived. Comprehensive simulation experiments, using measured acoustic transfer functions as well as real recordings on binaural hearing aids, demonstrate the capabilities of the BLCMV beamformer in various noise environments.**

***Index Terms*—Binaural cues, hearing aids, LCMV beamformer, noise reduction, relative transfer function.**

I. Introduction

THE OBJECTIVE of a binaural noise reduction algorithm is not only to selectively extract the desired speaker and to suppress interfering sources and ambient background noise, but also to preserve the auditory impression for the hearing aid user. On the one hand, for directional sources, preserving the auditory impression can be achieved by preserving the interaural time difference (ITD) and interaural level difference (ILD) cues of the sound sources in the acoustic scene. These binaural cues can be extracted from the so-called interaural transfer function (ITF), which is defined as the ratio of the acoustic transfer functions relating the source position and the two ears [1]. On the other hand, for ambient sources (e.g. diffuse noise), which cannot be properly described by the ITF, the diffuseness can be described by the coherence between both sides of the ears, i.e. the so-called interaural coherence (IC). Preserving the diffuseness of ambient sources is known to produce more natural sounds [2].

Many binaural noise reduction algorithms have been proposed that aim to preserve the binaural cues of the sound sources in the acoustic scene, which can be split into four main families. The first family is based on the concept of computational auditory scene analysis (CASA) [3]–[5], which aims to imitate the behavior of the human auditory system [6]. The second family consists of blind source separation (BSS) algorithms [7]–[9], which are based on the fundamental assumption of mutual statistical independence of the different source signals. The third family is based on a binaural version of the multichannel Wiener filter (MWF) [10]. The binaural MWF inherently preserves the binaural cues of the desired source but distorts the binaural cues of the noise (i.e. the beamformer imposes the noise to be coherent and perceived as arriving from the same direction as the desired source). Several extensions of the binaural MWF have been introduced aiming to also preserve the binaural cues of the noise [11]–[17]. By design, these methods suffer from some distortion of the desired source at the output. The fourth family is based on fixed or adaptive beamformers that aim to process the desired source without distortion [18]–[25]. Several minimum variance distortionless response (MVDR)-based beamformers with cue preservation capabilities can be found in [21], [23], [25]. In [20] a linearly constrained minimum variance (LCMV) criterion has been proposed in order to preserve the binaural cues of a single desired source by imposing multiple constraints. In order to preserve the binaural cues of the desired source, it is actually sufficient to preserve the so-called relative transfer function (RTF) of the desired source between the reference microphone signals on each hearing aid. In [24], the BLCMV beamformer was proposed, which imposes constraints on the LCMV cost function aiming to preserve the RTFs of both the desired and the interfering sources. Since the RTF is equivalent to the ITF for a binaural setup, the BLCMV is inherently capable of preserving the binaural cues of both the desired and the interfering sources, thus, making it particularly suitable for binaural hearing aid applications.

Our contribution in this paper is twofold. First, a theoretical analysis of the BLCMV beamformer is introduced. We propose several decompositions that reveal new insights into the performance of the BLCMV beamformer and its unique

Manuscript received July 30, 2015; revised December 13, 2015; accepted December 24, 2015. Date of publication January 05, 2016; date of current version February 23, 2016. This work was supported by the State of Lower Saxony, Hanover, Germany. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Richard Hendriks.

E. Hadad and S. Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan, 5290002, Israel (e-mail: Elior.Hadad@biu.ac.il; Sharon.Gannot@biu.ac.il).

S. Doclo is with the Department of Medical Physics and Acoustics and the Cluster of Excellence Hearing4All, University of Oldenburg, 26111 Oldenburg, Germany (e-mail: simon.doclo@uni-oldenburg.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASLP.2016.2514496

capabilities in terms of interference and noise reduction, while controlling the binaural cues of the desired and the interfering sources. Second, comprehensive simulation verification and experiments using measured acoustic transfer functions, as well as real recordings using binaural hearing aids, demonstrate the capabilities of this beamformer in various noise environments.

The paper is structured as follows: In Section II, the configuration and notation of the considered binaural hearing aid setup is introduced. In Section III, we review the BLCMV beamformer for the general case of multiple desired and multiple interfering sources, while depicting three proposed variants of the BLCMV criterion. We further show that the constraint sets of the BLCMV beamformer can be substituted by an equivalent modified constraint sets which can be estimated more easily. We focus on the dual source scenario. For this scenario, analytical expressions for the BLCMV beamformer are derived as well as three filter decompositions that provide some insights on the BLCMV beamformer capabilities. First, we show that the BLCMV beamformer can be decomposed into a sum of an LCMV beamformer reproducing the desired source while canceling the interfering source component, and an LCMV beamformer reproducing the interfering source component while canceling the desired source component. Second, an alternative decomposition of the BLCMV beamformer is derived as a weighted sum of two MVDR beamformers. Third, we show that the left and right filters of the BLCMV beamformer can be further interpreted as a common beamformer, followed by left and right *binauralization postfilters*, enabling to control the binaural cues for each constrained source. Since the left and right filters of the BLCMV beamformer, in general, are not parallel, this leads to the ability to preserve the binaural cues of both the desired and interfering sources and to impose the residual background noise to be non-coherent. The latter is advantageous compared to parallel beamformers, such as the binaural MWF and the binaural MVDR beamformers, for which the residual background noise is coherent and perceived as arriving from the desired source direction [10]. In Section IV, analytical expressions for the performance of the BLCMV beamformer are derived in terms of noise reduction, interference reduction and cue preservation. Moreover, various considerations are provided for setting the BLCMV parameters, allowing to control its performance. Section V is dedicated to describe the estimation procedure, restrictive assumptions regarding the activity of the sources, and the beamformer limitations. In Section VI, the theoretical results are validated by a comprehensive simulation verification and experiments using measured acoustic transfer functions as well as real recordings.

II. Problem Statement and Notation

A. Microphone and Output Signals

Consider an acoustic scenario consisting of desired and interfering sources in a noisy and reverberant environment. The sources are received by two fully connected hearing aid devices consisting of a microphone array with M_L microphones on the left hearing aid and M_R microphones on the right hearing aid, where $M = M_L + M_R$ denotes the total number of microphones as depicted in Fig. 1. The received

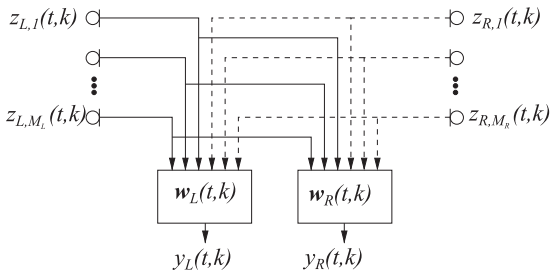

Fig. 1. General binaural processing scheme.

signal in the short-time Fourier transform (STFT) domain can be formulated as an M-dimensional vector $\boldsymbol{z}(t,k) = [z_{L,1}(t,k)\dots z_{L,M_L}(t,k) z_{R,1}(t,k)\dots z_{R,M_R}(t,k)]^T$, which can be written as

$$\begin{aligned}\boldsymbol{z}(t,k) &= \boldsymbol{z}_X(t,k) + \boldsymbol{z}_U(t,k) + \boldsymbol{z}_N(t,k)\\ &= \boldsymbol{z}_X(t,k) + \boldsymbol{z}_V(t,k),\end{aligned} \tag{1}$$

where k denotes the frequency index and t the frame index, and $\boldsymbol{z}_X(t,k)$, $\boldsymbol{z}_U(t,k)$, and $\boldsymbol{z}_N(t,k)$ denote the received desired source component, the received directional interfering (undesired) source component, and the received background noise component, respectively. $\boldsymbol{z}_V(t,k) = \boldsymbol{z}_U(t,k) + \boldsymbol{z}_N(t,k)$ is defined as the overall noise component as received by the microphones, i.e. the directional interfering source component plus the background noise component. The spatial correlation matrices of the desired source, interfering source and background noise components $\boldsymbol{R}_X$, $\boldsymbol{R}_U$ and $\boldsymbol{R}_N$, are defined as

$$\begin{aligned}\boldsymbol{R}_X(t,k) &= \mathcal{E}\{\boldsymbol{z}_X(t,k)\boldsymbol{z}_X^H(t,k)\},\\ \boldsymbol{R}_U(t,k) &= \mathcal{E}\{\boldsymbol{z}_U(t,k)\boldsymbol{z}_U^H(t,k)\},\\ \boldsymbol{R}_N(t,k) &= \mathcal{E}\{\boldsymbol{z}_N(t,k)\boldsymbol{z}_N^H(t,k)\},\end{aligned} \tag{2}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator. Assuming statistical independence between the components in (1), the spatial correlation matrix of the microphone signals $\boldsymbol{R}_Z$ can be written as

$$\boldsymbol{R}_Z(t,k) = \boldsymbol{R}_X(t,k) + \underbrace{\boldsymbol{R}_U(t,k) + \boldsymbol{R}_N(t,k)}_{\boldsymbol{R}_V(t,k)}, \tag{3}$$

with $\boldsymbol{R}_V$ the spatial correlation matrix of the overall noise component.

Let m_L and m_R be the indices of the left and right reference microphones, respectively (usually selected as the microphones closest to the ears). The respective reference microphone signals at the left and the right hearing aids are given by

$$z_L(t,k) = \boldsymbol{e}_L^H \boldsymbol{z}(t,k), \quad z_R(t,k) = \boldsymbol{e}_R^H \boldsymbol{z}(t,k), \tag{4}$$

where $\boldsymbol{e}_L$ and $\boldsymbol{e}_R$ are M-dimensional vectors with '1' in the m_Lth and m_Rth component, respectively, and '0' elsewhere. Two spatial M-dimensional filters $\boldsymbol{w}_L(t,k)$ and $\boldsymbol{w}_R(t,k)$ (one for each side), utilizing all M microphones, constitute the binaural beamformer, i.e.

$$y_L(t,k) = \boldsymbol{w}_L^H \boldsymbol{z}(t,k), \quad y_R(t,k) = \boldsymbol{w}_R^H \boldsymbol{z}(t,k). \tag{5}$$

Henceforth, t and k will be omitted for the sake of brevity.

B. Interaural Criteria

The ITF is defined as the ratio of the components at the reference microphones at both hearing aids. For practical implementations, the input and output ITFs of the desired source can be estimated from the spatial correlation matrix [12], i.e.

$$\mathrm{ITF}_{X,\mathrm{IN}} = \frac{\boldsymbol{e}_L^H \boldsymbol{R}_X \boldsymbol{e}_L}{\boldsymbol{e}_R^H \boldsymbol{R}_X \boldsymbol{e}_L}, \quad \mathrm{ITF}_{X,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{R}_X \boldsymbol{w}_L}{\boldsymbol{w}_R^H \boldsymbol{R}_X \boldsymbol{w}_L}. \quad (6)$$

The ITF is a complex-valued frequency-dependent scalar, from which the ILD and the ITD binaural cues can be computed as [12]

$$\mathrm{ILD} = 20\log_{10}(|\mathrm{ITF}|), \quad \mathrm{ITD} = \frac{\angle(\mathrm{ITF})}{\omega}, \quad (7)$$

with $\angle$ denoting the unwrapped phase and ω the radian frequency. Note that (6)–(7) will also be used in the general case where the rank of $\boldsymbol{R}_X$ is larger than one. The ITF of the interfering source and the background noise can be estimated in a similar way.

As explained above, for diffuse noise, the perceptual impression can not be properly described by the ITF, but rather by the IC. The input and output ICs are given by

$$\mathrm{IC}_{\mathrm{IN}} = \frac{\mathcal{E}\{z_L z_R^*\}}{\sqrt{\mathcal{E}\{z_L z_L^*\}}\sqrt{\mathcal{E}\{z_R z_R^*\}}},$$
$$\mathrm{IC}_{\mathrm{OUT}} = \frac{\mathcal{E}\{y_L {y_R}^*\}}{\sqrt{\mathcal{E}\{y_L y_L^*\}}\sqrt{\mathcal{E}\{y_R {y_R}^*\}}}, \quad (8)$$

and the real-valued magnitude squared coherence (MSC) is defined as

$$\mathrm{MSC}_{\mathrm{IN}} = |\mathrm{IC}_{\mathrm{IN}}|^2, \quad \mathrm{MSC}_{\mathrm{OUT}} = |\mathrm{IC}_{\mathrm{OUT}}|^2. \quad (9)$$

The input and output ICs of the background noise can be estimated from the spatial correlation matrix and are given by [14]

$$\mathrm{IC}_{N,\mathrm{IN}} = \frac{\boldsymbol{e}_L^H \boldsymbol{R}_N \boldsymbol{e}_R}{\sqrt{\boldsymbol{e}_L^H \boldsymbol{R}_N \boldsymbol{e}_L}\sqrt{\boldsymbol{e}_R^H \boldsymbol{R}_N \boldsymbol{e}_R}},$$
$$\mathrm{IC}_{N,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{R}_N \boldsymbol{w}_R}{\sqrt{\boldsymbol{w}_L^H \boldsymbol{R}_N \boldsymbol{w}_L}\sqrt{\boldsymbol{w}_R^H \boldsymbol{R}_N \boldsymbol{w}_R}}. \quad (10)$$

C. Power Spectral Density

The input and output power spectral densities (PSDs) of the desired source component for the left and right filters are given by

$$S_{X,L,\mathrm{IN}} = \boldsymbol{e}_L^H \boldsymbol{R}_X \boldsymbol{e}_L, \quad S_{X,L,\mathrm{OUT}} = \boldsymbol{w}_L^H \boldsymbol{R}_X \boldsymbol{w}_L,$$
$$S_{X,R,\mathrm{IN}} = \boldsymbol{e}_R^H \boldsymbol{R}_X \boldsymbol{e}_R, \quad S_{X,R,\mathrm{OUT}} = \boldsymbol{w}_R^H \boldsymbol{R}_X \boldsymbol{w}_R. \quad (11)$$

The input and output PSDs of the received signal, the interfering source, the overall noise, and the background noise components are defined similarly by substituting $\boldsymbol{R}_X$ in (11) with $\boldsymbol{R}_Z$, $\boldsymbol{R}_U$, $\boldsymbol{R}_V$, or $\boldsymbol{R}_N$, respectively.

D. Dual Source Scenario

In this section, we consider a common dual source (DS) scenario, consisting one desired source, one interfering source (e.g. competing speakers) and background noise, which can be either directional, non-directional or a combination. The desired and the interfering source components can be modeled as

$$\boldsymbol{z}_X = s_X \boldsymbol{a}, \quad \boldsymbol{z}_U = s_U \boldsymbol{b}, \quad (12)$$

where the M-dimensional vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ denote the acoustic transfer functions (ATFs) from the sources to the microphones, and s_X and s_U denote the desired and interfering source signals, respectively. In this case, the correlation matrices $\boldsymbol{R}_X$ and $\boldsymbol{R}_U$ are rank-1 matrices, i.e.

$$\boldsymbol{R}_X = P_S \boldsymbol{a}\boldsymbol{a}^H, \quad \boldsymbol{R}_U = P_U \boldsymbol{b}\boldsymbol{b}^H, \quad (13)$$

with $P_S = \mathcal{E}\{|s_X|^2\}$ and $P_U = \mathcal{E}\{|s_U|^2\}$ denoting the PSD of the desired and interfering source components, respectively. Using (6), the input ITFs of the desired and interfering sources are equal to

$$\mathrm{ITF}_{X,\mathrm{IN}} = \frac{\boldsymbol{e}_L^H P_S \boldsymbol{a}\boldsymbol{a}^H \boldsymbol{e}_L}{\boldsymbol{e}_R^H P_S \boldsymbol{a}\boldsymbol{a}^H \boldsymbol{e}_L} = \frac{a_L a_L^*}{a_R a_L^*} = \frac{a_L}{a_R},$$
$$\mathrm{ITF}_{U,\mathrm{IN}} = \frac{\boldsymbol{e}_L^H P_U \boldsymbol{b}\boldsymbol{b}^H \boldsymbol{e}_L}{\boldsymbol{e}_R^H P_U \boldsymbol{b}\boldsymbol{b}^H \boldsymbol{e}_L} = \frac{b_L b_L^*}{b_R b_L^*} = \frac{b_L}{b_R}. \quad (14)$$

Note that for directional sources, the IC is equal to the normalized ITF [14], and can be solely described by the ITD, i.e.

$$\mathrm{IC} = \frac{\mathrm{ITF}}{|\mathrm{ITF}|} = e^{j\omega \cdot \mathrm{ITD}}. \quad (15)$$

This implies that directional sources are characterised by an MSC equal to '1', and can be fully described by the ITF. On the other hand, diffuse sources are fully described by the IC. The spatial perception of a source is based on a combination of the ITF and the IC, where the IC determines the reliability of the ILD and ITD cues extracted from the ITF [26], [27].

E. Multiple Sources Scenario

In this section, we consider the general acoustic scenario with multiple desired and multiple interfering sources (e.g. simultaneous conservations). Consider N_X desired sources $s_X^1, \ldots, s_X^{N_X}$ and N_U interfering sources $s_U^1, \ldots, s_U^{N_U}$. The received signal vector can be written as

$$\boldsymbol{z} = \sum_{m=1}^{N_X} \boldsymbol{z}_X^m + \sum_{m=1}^{N_U} \boldsymbol{z}_U^m + \boldsymbol{z}_N = \boldsymbol{A} \cdot \boldsymbol{s}_X + \boldsymbol{B} \cdot \boldsymbol{s}_U + \boldsymbol{z}_N, \quad (16)$$

where $\boldsymbol{z}_X^m$ and $\boldsymbol{z}_U^m$ denote the mth desired and mth interfering source components, respectively, and $\boldsymbol{s}_X = [s_X^1 \ \cdots \ s_X^{N_X}]^T$ and $\boldsymbol{s}_U = [s_U^1 \ \cdots \ s_U^{N_U}]^T$ are vectors comprising the desired and interfering source signals, respectively. $\boldsymbol{A} = [\boldsymbol{a}^1 \ \cdots \ \boldsymbol{a}^{N_X}]$ and $\boldsymbol{B} = [\boldsymbol{b}^1 \ \cdots \ \boldsymbol{b}^{N_U}]$ are $M \times N_X$-dimensional and $M \times N_U$-dimensional matrices, respectively, comprising of

concatenated ATFs relating the desired and interfering sources to the microphones. The spatial correlation matrix of the microphone signals $\boldsymbol{R}_Z$ can now be written as

$$\boldsymbol{R}_Z = \underbrace{\boldsymbol{A}\boldsymbol{\Lambda}_X\boldsymbol{A}^H}_{\boldsymbol{R}_X} + \underbrace{\boldsymbol{B}\boldsymbol{\Lambda}_U\boldsymbol{B}^H}_{\boldsymbol{R}_U} + \boldsymbol{R}_N, \tag{17}$$

where $\boldsymbol{\Lambda}_X \triangleq \mathrm{diag}([P_S^1 \ \dots \ P_S^{N_X}])$ and $\boldsymbol{\Lambda}_U \triangleq \mathrm{diag}([P_U^1 \ \dots \ P_U^{N_U}])$ are diagonal matrices containing the PSDs of the desired and interfering sources on their main diagonal, respectively.

III. Binaural Noise Reduction Technique

In this section, the proposed algorithm is derived. In Section III-A, we review the BLCMV beamformer for the general case of multiple desired and multiple interfering sources. In Section III-B, we discuss the variants of the BLCMV criterion. In Section III-C, we show that the constraint sets can be substituted by an equivalent modified constraint sets which can be estimated more easily. In Sections III-D to III-G, for simplicity, we focus on the dual source scenario. For this scenario, analytical expressions for the BLCMV beamformer are derived as well as three filter decompositions that provide some insights into the BLCMV beamformer capabilities.

A. The Binaural LCMV Beamformer (BLCMV)

The BLCMV beamformer proposed in [24] consists of two filters designed to reproduce a filtered version of the desired source component as received by the reference microphones in each hearing aid, while reducing the interfering source component and minimizing the *output power*.

Consider a general acoustic scenario with N_X desired sources and N_U interfering sources, as described in Section II-E. The BLCMV beamformer is constructed using two sets of linear constraints. One set is imposing constraints on the desired sources, i.e.

$$\boldsymbol{w}_L^H\boldsymbol{A} = \xi\boldsymbol{e}_L^H\boldsymbol{A}, \quad \boldsymbol{w}_R^H\boldsymbol{A} = \xi\boldsymbol{e}_R^H\boldsymbol{A}, \tag{18}$$

while another set is imposing constraints on the interfering sources, i.e.

$$\boldsymbol{w}_L^H\boldsymbol{B} = \eta\boldsymbol{e}_L^H\boldsymbol{B}, \quad \boldsymbol{w}_R^H\boldsymbol{B} = \eta\boldsymbol{e}_R^H\boldsymbol{B}, \tag{19}$$

where $0 < \xi \leq 1$ and $0 \leq \eta \leq 1$ are real-valued scalars defined as the *cue gain factors* for the desired and the interfering sources, respectively. Typically ξ will be close to '1' (limiting the distortion for the desired sources), whereas η will be smaller than '1' (suppressing the interfering sources). These constraint sets can be combined to define a general LCMV criterion with multiple constraints on both the desired and the interfering sources for the left and right filters minimizing the *output power*:

$$\begin{aligned}\boldsymbol{w}_L &= \arg\min_{\boldsymbol{w}_L}\{S_{Z,L,\mathrm{OUT}}\} \text{ s.t. } \boldsymbol{C}^H\boldsymbol{w}_L = \boldsymbol{g}_L,\\ \boldsymbol{w}_R &= \arg\min_{\boldsymbol{w}_R}\{S_{Z,R,\mathrm{OUT}}\} \text{ s.t. } \boldsymbol{C}^H\boldsymbol{w}_R = \boldsymbol{g}_R,\end{aligned} \tag{20}$$

where $S_{Z,L,\mathrm{OUT}} = \boldsymbol{w}_L^H\boldsymbol{R}_Z\boldsymbol{w}_L$, $S_{Z,R,\mathrm{OUT}} = \boldsymbol{w}_R^H\boldsymbol{R}_Z\boldsymbol{w}_R$, and the left and the right constraint sets are given by

$$\boldsymbol{C}^H\boldsymbol{w}_L = \boldsymbol{g}_L, \quad \boldsymbol{C}^H\boldsymbol{w}_R = \boldsymbol{g}_R. \tag{21}$$

The left and right response vectors are defined as

$$\boldsymbol{g}_L = \left[\xi\boldsymbol{e}_L^H\boldsymbol{A} \ \eta\boldsymbol{e}_L^H\boldsymbol{B}\right]^H, \boldsymbol{g}_R = \left[\xi\boldsymbol{e}_R^H\boldsymbol{A} \ \eta\boldsymbol{e}_R^H\boldsymbol{B}\right]^H, \tag{22}$$

and the constraint matrix is defined as

$$\boldsymbol{C} = \left[\boldsymbol{A} \ \boldsymbol{B}\right]. \tag{23}$$

Note that the constraint sets in (21) utilizing the ATFs can be reformulated as constraint sets utilizing the RTFs, i.e.

$$\tilde{\boldsymbol{C}}_L^H\boldsymbol{w}_L = \tilde{\boldsymbol{g}}_L, \quad \tilde{\boldsymbol{C}}_R^H\boldsymbol{w}_R = \tilde{\boldsymbol{g}}_R, \tag{24}$$

with the left and right RTF response vectors defined as

$$\tilde{\boldsymbol{g}}_L = \tilde{\boldsymbol{g}}_R = \left[\xi\boldsymbol{1}_{1\times N_X} \ \eta\boldsymbol{1}_{1\times N_U}\right]^H, \tag{25}$$

with $\boldsymbol{1}_{1\times K}$ denoting the K-dimensional row vector of all ones. The left and right RTF constraint matrices are defined as

$$\tilde{\boldsymbol{C}}_L = \left[\tilde{\boldsymbol{A}}_L \quad \tilde{\boldsymbol{B}}_L\right], \quad \tilde{\boldsymbol{C}}_R = \left[\tilde{\boldsymbol{A}}_R \quad \tilde{\boldsymbol{B}}_R\right], \tag{26}$$

where $\tilde{\boldsymbol{A}}_L = [\frac{\boldsymbol{a}^1}{a_L^1} \ \cdots \ \frac{\boldsymbol{a}^{N_X}}{a_L^{N_X}}]$ and $\tilde{\boldsymbol{B}}_L = [\frac{\boldsymbol{b}^1}{b_L^1} \ \cdots \ \frac{\boldsymbol{b}^{N_U}}{b_L^{N_U}}]$ are $M \times N_X$-dimensional and $M \times N_U$-dimensional matrices, respectively, comprising concatenated ATFs relating the desired and interfering sources and the microphones normalized by the left reference microphone. $\tilde{\boldsymbol{A}}_R$ and $\tilde{\boldsymbol{B}}_R$ are defined similarly for the right side. For the sake of readability, note that all derivations in the paper will use the definitions (22) and (23) based on ATFs, while in practice (25) and (26), based on RTFs, will be used.

The well-known closed-form solution of the left and right filters of the BLCMV beamformer in (20) is given by [28]

$$\begin{aligned}\boldsymbol{w}_L &= \boldsymbol{R}_Z^{-1}\boldsymbol{C}\left[\boldsymbol{C}^H\boldsymbol{R}_Z^{-1}\boldsymbol{C}\right]^{-1}\boldsymbol{g}_L,\\ \boldsymbol{w}_R &= \boldsymbol{R}_Z^{-1}\boldsymbol{C}\left[\boldsymbol{C}^H\boldsymbol{R}_Z^{-1}\boldsymbol{C}\right]^{-1}\boldsymbol{g}_R.\end{aligned} \tag{27}$$

The proposed BLCMV beamformer has several advantages. First, the simple direct path model is generalized by the ATFs relating the sources and the microphones [29]. Constructing the LCMV beamformer using the ATFs circumvents the self-cancelation phenomenon that is frequently encountered when using the simple direct path model. Second, while estimating the ATFs is a cumbersome task, practical estimation procedures for estimating the RTF exist (cf. Section V and [29]–[31]). Third, in binaural signal processing algorithms, it is desirable to preserve the binaural cues of both the desired and interfering sources. As the BLCMV beamformer preserves the binaural cues for all constrained sources, this capability is a major advantage in binaural hearing aid applications. Fourth, the cue gain factors in (18) and (19) enable to control the level of speech distortion, interference reduction and noise reduction. Using real-valued and frequency-independent cue gain factors ensures that no amplitude and phase distortion is imposed on the constrained sources.

B. Variants of the BLCMV Beamformer

Instead of minimizing the output power, two other variants of the proposed beamformer can be derived. The BLCMV beamformer minimizing the *overall noise power* subject to the constraint in (21) is given by

$$\begin{aligned}\boldsymbol{w}_L &= \arg\min_{\boldsymbol{w}_L}\{S_{V,L,\mathrm{OUT}}\}\ \text{s.t.}\ \boldsymbol{C}^H\boldsymbol{w}_L = \boldsymbol{g}_L,\\ \boldsymbol{w}_R &= \arg\min_{\boldsymbol{w}_R}\{S_{V,R,\mathrm{OUT}}\}\ \text{s.t.}\ \boldsymbol{C}^H\boldsymbol{w}_R = \boldsymbol{g}_R,\end{aligned} \tag{28}$$

where $S_{V,L,\mathrm{OUT}} = \boldsymbol{w}_L^H\boldsymbol{R}_V\boldsymbol{w}_L$, $S_{V,R,\mathrm{OUT}} = \boldsymbol{w}_R^H\boldsymbol{R}_V\boldsymbol{w}_R$, and the closed-form solution of the left and right filters is given by

$$\begin{aligned}\boldsymbol{w}_L &= \boldsymbol{R}_V^{-1}\boldsymbol{C}\left[\boldsymbol{C}^H\boldsymbol{R}_V^{-1}\boldsymbol{C}\right]^{-1}\boldsymbol{g}_L,\\ \boldsymbol{w}_R &= \boldsymbol{R}_V^{-1}\boldsymbol{C}\left[\boldsymbol{C}^H\boldsymbol{R}_V^{-1}\boldsymbol{C}\right]^{-1}\boldsymbol{g}_R.\end{aligned} \tag{29}$$

As a result of the constraints on the desired sources in (18), using (17) and (11), the output PSD of the desired source components of the BLCMV beamformer is independent of the minimization criterion, i.e.

$$\begin{aligned}S_{X,L,\mathrm{OUT}} &= \xi^2\boldsymbol{e}_L^H\boldsymbol{A}\boldsymbol{\Lambda}_X\boldsymbol{A}^H\boldsymbol{e}_L = \xi^2\boldsymbol{e}_L^H\boldsymbol{R}_X\boldsymbol{e}_L,\\ S_{X,R,\mathrm{OUT}} &= \xi^2\boldsymbol{e}_R^H\boldsymbol{A}\boldsymbol{\Lambda}_X\boldsymbol{A}^H\boldsymbol{e}_R = \xi^2\boldsymbol{e}_R^H\boldsymbol{R}_X\boldsymbol{e}_R.\end{aligned} \tag{30}$$

Since $S_{Z,L,\mathrm{OUT}} = S_{X,L,\mathrm{OUT}} + S_{V,L,\mathrm{OUT}}$ and $S_{Z,R,\mathrm{OUT}} = S_{X,R,\mathrm{OUT}} + S_{V,R,\mathrm{OUT}}$, this implies that ideally, i.e. without RTF estimation errors of the desire source, both constrained optimization problems in (20) and (28) yield the same solutions [32].

Alternatively, the BLCMV beamformer that minimizing the *background noise power* subject to the constraint in (21) is given by

$$\begin{aligned}\boldsymbol{w}_L &= \arg\min_{\boldsymbol{w}_L}\{S_{N,L,\mathrm{OUT}}\}\ \text{s.t.}\ \boldsymbol{C}^H\boldsymbol{w}_L = \boldsymbol{g}_L,\\ \boldsymbol{w}_R &= \arg\min_{\boldsymbol{w}_R}\{S_{N,R,\mathrm{OUT}}\}\ \text{s.t.}\ \boldsymbol{C}^H\boldsymbol{w}_R = \boldsymbol{g}_R,\end{aligned} \tag{31}$$

where $S_{N,L,\mathrm{OUT}} = \boldsymbol{w}_L^H\boldsymbol{R}_N\boldsymbol{w}_L$, $S_{N,R,\mathrm{OUT}} = \boldsymbol{w}_R^H\boldsymbol{R}_N\boldsymbol{w}_R$, and the closed-form solution of the left and right filters is given by

$$\begin{aligned}\boldsymbol{w}_L &= \boldsymbol{R}_N^{-1}\boldsymbol{C}\left[\boldsymbol{C}^H\boldsymbol{R}_N^{-1}\boldsymbol{C}\right]^{-1}\boldsymbol{g}_L,\\ \boldsymbol{w}_R &= \boldsymbol{R}_N^{-1}\boldsymbol{C}\left[\boldsymbol{C}^H\boldsymbol{R}_N^{-1}\boldsymbol{C}\right]^{-1}\boldsymbol{g}_R.\end{aligned} \tag{32}$$

As a result of the constraints on the interfering sources in (19), using (17), the output PSD of the interfering source components of the BLCMV beamformer is independent of the minimization criterion, i.e.

$$\begin{aligned}S_{U,L,\mathrm{OUT}} &= \eta^2\boldsymbol{e}_L^H\boldsymbol{B}\boldsymbol{\Lambda}_U\boldsymbol{B}^H\boldsymbol{e}_L = \eta^2\boldsymbol{e}_L^H\boldsymbol{R}_U\boldsymbol{e}_L,\\ S_{U,R,\mathrm{OUT}} &= \eta^2\boldsymbol{e}_R^H\boldsymbol{B}\boldsymbol{\Lambda}_U\boldsymbol{B}^H\boldsymbol{e}_R = \eta^2\boldsymbol{e}_R^H\boldsymbol{R}_U\boldsymbol{e}_R.\end{aligned} \tag{33}$$

Since $S_{V,L,\mathrm{OUT}} = S_{U,L,\mathrm{OUT}} + S_{N,L,\mathrm{OUT}}$ and $S_{V,R,\mathrm{OUT}} = S_{U,R,\mathrm{OUT}} + S_{N,R,\mathrm{OUT}}$, this implies that ideally, i.e. without RTF estimation errors of the interfering source, both constrained optimization problems in (28) and (31) yield the same solutions as well.

In this paper we will adopt the BLCMV variant using $\boldsymbol{R}_N$ in (31) for two reasons. First, assuming that the background noise is stationary, the spatial correlation matrix $\boldsymbol{R}_N$ can be estimated more easily than the highly time-varying spatial correlation matrices $\boldsymbol{R}_Z$ and $\boldsymbol{R}_V$ (cf. Section V). Second, there is a difference in the robustness to RTF estimation errors for the three considered variants of the BLCMV beamformer. In [33], [34], two variants of the (monaural) MVDR beamformer were compared, where it was shown that the MVDR beamformer using $\boldsymbol{R}_V$ is more robust to steering vector errors (of the desired source) than the MVDR beamformer using $\boldsymbol{R}_Z$. Similarly to the MVDR beamformer, we postulate that the BLCMV beamformer using $\boldsymbol{R}_N$ is more robust to steering vector errors (of both the desired and the interfering sources) than the other BLCMV variants. Particularly, for the BLCMV beamformer using $\boldsymbol{R}_Z$, RTF estimation errors of the desired source may result in a suppression of the desired source component at the output of the beamformer. Similarly, for the BLCMV beamformer using $\boldsymbol{R}_V$, RTF estimation errors of the interfering source may result in a suppression of the interfering source component at the output of the beamformer to a level that is lower than η. On the contrary, RTF estimation errors of both the desired and interfering sources are excluded from the minimization of the BLCMV criterion using $\boldsymbol{R}_N$. A theoretical analysis of the robustness of the BLCMV variants to RTF estimation errors (similarly to the robustness analysis of the monaural LCMV beamformer in [32], [33], [35]) remains a topic for future research.

C. Subspace Constraint Sets

For implementing the constraint sets in (24) an estimate of the RTFs of the desired and interfering sources is required. Obtaining such estimates might be a cumbersome task in practical scenarios, since it is usually required that the sources are not active simultaneously. In this section, we prove that substituting the individual RTFs of the desired and interfering sources in the constraint sets by the corresponding basis vectors that span the set of RTFs, results in an equivalent beamformer. These basis vectors can be more easily estimated, as described in [31], [36] and discussed in Section V.

Denote by $\boldsymbol{Q}_X \triangleq [\boldsymbol{q}_X^1\ \ldots\ \boldsymbol{q}_X^{N_X}]$ a basis spanning the subspace of ATFs of the desired sources $\boldsymbol{A}$ and by $\boldsymbol{Q}_U \triangleq [\boldsymbol{q}_U^1\ \ldots\ \boldsymbol{q}_U^{N_U}]$ a basis spanning the subspace of ATFs of the interfering sources $\boldsymbol{B}$, i.e. $\boldsymbol{A} = \boldsymbol{Q}_X\boldsymbol{\Theta}_X$ and $\boldsymbol{B} = \boldsymbol{Q}_U\boldsymbol{\Theta}_U$ where $\boldsymbol{\Theta}_X \triangleq [\theta_X^1\ \ldots\ \theta_X^{N_X}]$ and $\boldsymbol{\Theta}_U \triangleq [\boldsymbol{\theta}_U^1\ \ldots\ \boldsymbol{\theta}_U^{N_U}]$ are the projection coefficients matrices. Using both subspaces, define the subspace constraint matrix as

$$\breve{\boldsymbol{C}} = \left[\boldsymbol{Q}_X\ \boldsymbol{Q}_U\right], \tag{34}$$

and define the left and right subspace response vectors as

$$\begin{aligned}\breve{\boldsymbol{g}}_L &= \left[\xi\boldsymbol{e}_L^H\boldsymbol{Q}_X\ \ \eta\boldsymbol{e}_L^H\boldsymbol{Q}_U\right]^H,\\ \breve{\boldsymbol{g}}_R &= \left[\xi\boldsymbol{e}_R^H\boldsymbol{Q}_X\ \ \eta\boldsymbol{e}_R^H\boldsymbol{Q}_U\right]^H.\end{aligned} \tag{35}$$

The subspace constraint sets are then defined as

$$\breve{\boldsymbol{C}}^H\boldsymbol{w}_L = \breve{\boldsymbol{g}}_L,\quad \breve{\boldsymbol{C}}^H\boldsymbol{w}_R = \breve{\boldsymbol{g}}_R. \tag{36}$$

We prove now that satisfying the subspace constraint sets in (36) is equivalent to satisfying the original constraint sets in (18) and (19). By applying the left and right filters of the BLCMV beamformer to the mth received desired source component, we obtain

$$\begin{aligned}\boldsymbol{w}_L^H \boldsymbol{z}_X^m &= \boldsymbol{w}_L^H s_X^m \boldsymbol{a}^m = s_X^m \boldsymbol{w}_L^H \boldsymbol{Q}_X \boldsymbol{\theta}_X^m,\\ \boldsymbol{w}_R^H \boldsymbol{z}_X^m &= \boldsymbol{w}_R^H s_X^m \boldsymbol{a}^m = s_X^m \boldsymbol{w}_R^H \boldsymbol{Q}_X \boldsymbol{\theta}_X^m,\end{aligned} \tag{37}$$

where, by the basis construction, $\boldsymbol{a}^m = \boldsymbol{Q}_X \boldsymbol{\theta}_X^m$. In addition, from (36), we obtain

$$\boldsymbol{w}_L^H \boldsymbol{Q}_X = \xi \boldsymbol{e}_L^H \boldsymbol{Q}_X, \boldsymbol{w}_R^H \boldsymbol{Q}_X = \xi \boldsymbol{e}_R^H \boldsymbol{Q}_X. \tag{38}$$

Substituting (38) into (37) yields

$$\begin{aligned}\boldsymbol{w}_L^H \boldsymbol{z}_X^m &= s_X^m \xi \boldsymbol{e}_L^H \boldsymbol{Q}_X \boldsymbol{\theta}_X^m = s_X^m \xi \boldsymbol{e}_L^H \boldsymbol{a}^m = \xi s_X^m a_L^m,\\ \boldsymbol{w}_R^H \boldsymbol{z}_X^m &= s_X^m \xi \boldsymbol{e}_R^H \boldsymbol{Q}_X \boldsymbol{\theta}_X^m = s_X^m \xi \boldsymbol{e}_R^H \boldsymbol{a}^m = \xi s_X^m a_R^m.\end{aligned} \tag{39}$$

Since $\boldsymbol{z}_X^m = s_X^m \boldsymbol{a}^m$ we can deduce from (39) that

$$\boldsymbol{w}_L^H \boldsymbol{a}^m = \xi a_L^m, \boldsymbol{w}_R^H \boldsymbol{a}^m = \xi a_R^m, \tag{40}$$

which is identical to the original constraint set in (18). A similar derivation applies to the interfering sources.

D. The BLCMV Beamformer for the Dual Source Scenario

In this section, we focus on the DS scenario discussed in Section II-D, corresponding to a single constraint on the desired source response and a single constraint on the interfering source response. Hence, the constraint matrix in (23) is simplified to

$$\boldsymbol{C} = [\boldsymbol{a} \quad \boldsymbol{b}], \tag{41}$$

and the corresponding left and right response vectors in (22) are given by

$$\boldsymbol{g}_L = \begin{bmatrix}\xi a_L & \eta b_L\end{bmatrix}^H, \quad \boldsymbol{g}_R = \begin{bmatrix}\xi a_R & \eta b_R\end{bmatrix}^H. \tag{42}$$

In the next sections, we derive several decompositions for the filters. Similarly to [33], we define generalized inner products between the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$, i.e.

$$\begin{aligned}\gamma_a &= \boldsymbol{a}^H \boldsymbol{R}_N^{-1} \boldsymbol{a} = \bar{\boldsymbol{a}}^H \bar{\boldsymbol{a}},\\ \gamma_b &= \boldsymbol{b}^H \boldsymbol{R}_N^{-1} \boldsymbol{b} = \bar{\boldsymbol{b}}^H \bar{\boldsymbol{b}},\\ \gamma_{ab} &= \boldsymbol{a}^H \boldsymbol{R}_N^{-1} \boldsymbol{b} = \bar{\boldsymbol{a}}^H \bar{\boldsymbol{b}},\end{aligned} \tag{43}$$

with $\bar{\boldsymbol{a}} = \boldsymbol{R}_N^{-\frac{1}{2}} \boldsymbol{a}$ and $\bar{\boldsymbol{b}} = \boldsymbol{R}_N^{-\frac{1}{2}} \boldsymbol{b}$ denoting the noise-prewhitened ATFs of the desired and interfering sources[1]. It is assumed that $\boldsymbol{R}_N$, and consequently $\boldsymbol{R}_N^{-1}$, are positive definite Hermitian matrices. The cosine-squared of the generalized angle between $\boldsymbol{a}$ and $\boldsymbol{b}$ is given by

$$\Gamma = \cos^2(\boldsymbol{\theta}_{ab}) = \frac{|\bar{\boldsymbol{a}}^H \bar{\boldsymbol{b}}|^2}{\|\bar{\boldsymbol{a}}\|^2 \|\bar{\boldsymbol{b}}\|^2} = \frac{|\gamma_{ab}|^2}{\gamma_a \gamma_b}, \tag{44}$$

with $\boldsymbol{\theta}_{ab}$ the angle between $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$. Using the Cauchy-Schwarz inequality it can be shown that $0 \leq \Gamma \leq 1$. Other forms of prewhitening, e.g. using $\boldsymbol{R}_N^{-\alpha}$ for various values of α, and its properties can be found in [37], [38].

[1]Note that $\boldsymbol{a}$ and $\boldsymbol{b}$ are the entire ATFs of the sources, including reverberation, such that this definition generalizes the definition in [33], which considers steering vectors in free-space propagation.

E. Filter Decomposition into Two BLCMV Beamformers

Substituting (42) into (32), the left and right filters of the BLCMV beamformer can be written as a sum of two beamformers, i.e.

$$\boldsymbol{w}_L = \boldsymbol{w}_{X,L} + \boldsymbol{w}_{U,L}, \quad \boldsymbol{w}_R = \boldsymbol{w}_{X,R} + \boldsymbol{w}_{U,R}, \tag{45}$$

with

$$\begin{aligned}\boldsymbol{w}_{X,L} &= \boldsymbol{R}_N^{-1} \boldsymbol{C} \left[\boldsymbol{C}^H \boldsymbol{R}_N^{-1} \boldsymbol{C}\right]^{-1} \boldsymbol{g}_{X,L},\\ \boldsymbol{w}_{U,L} &= \boldsymbol{R}_N^{-1} \boldsymbol{C} \left[\boldsymbol{C}^H \boldsymbol{R}_N^{-1} \boldsymbol{C}\right]^{-1} \boldsymbol{g}_{U,L},\\ \boldsymbol{w}_{X,R} &= \boldsymbol{R}_N^{-1} \boldsymbol{C} \left[\boldsymbol{C}^H \boldsymbol{R}_N^{-1} \boldsymbol{C}\right]^{-1} \boldsymbol{g}_{X,R},\\ \boldsymbol{w}_{U,R} &= \boldsymbol{R}_N^{-1} \boldsymbol{C} \left[\boldsymbol{C}^H \boldsymbol{R}_N^{-1} \boldsymbol{C}\right]^{-1} \boldsymbol{g}_{U,R},\end{aligned} \tag{46}$$

and

$$\begin{aligned}\boldsymbol{g}_{X,L} &= \begin{bmatrix}\xi a_L & 0\end{bmatrix}^H, \quad \boldsymbol{g}_{U,L} = \begin{bmatrix}0 & \eta b_L\end{bmatrix}^H,\\ \boldsymbol{g}_{X,R} &= \begin{bmatrix}\xi a_R & 0\end{bmatrix}^H, \quad \boldsymbol{g}_{U,R} = \begin{bmatrix}0 & \eta b_R\end{bmatrix}^H,\end{aligned} \tag{47}$$

where $\boldsymbol{g}_L = \boldsymbol{g}_{X,L} + \boldsymbol{g}_{U,L}$ and $\boldsymbol{g}_R = \boldsymbol{g}_{X,R} + \boldsymbol{g}_{U,R}$. $\boldsymbol{w}_{X,L}$ and $\boldsymbol{w}_{X,R}$ are denoted as the left and right filters of the desired BLCMV beamformer (D-BLCMV), while $\boldsymbol{w}_{U,L}$ and $\boldsymbol{w}_{U,R}$ are denoted as the left and right filters of the undesired BLCMV beamformer (U-BLCMV) respectively. The left filter of the D-BLCMV beamformer reproduces the desired source component as received by the left reference microphone (multiplied by ξ), while entirely canceling the interfering source component and minimizing the background noise power, i.e.

$$\boldsymbol{w}_{X,L} = \underset{\boldsymbol{w}_{X,L}}{\arg\min} \{\boldsymbol{w}_{X,L}^H \boldsymbol{R}_N \boldsymbol{w}_{X,L}\} \text{ s.t. } \boldsymbol{C}^H \boldsymbol{w}_{X,L} = \boldsymbol{g}_{X,L}. \tag{48}$$

Similarly, the left filter of the U-BLCMV beamformer reproduces the interfering source component as received by the left reference microphone (multiplied by η), while entirely canceling the desired source component and minimizing the background noise power, i.e.

$$\boldsymbol{w}_{U,L} = \underset{\boldsymbol{w}_{U,L}}{\arg\min} \{\boldsymbol{w}_{U,L}^H \boldsymbol{R}_N \boldsymbol{w}_{U,L}\} \text{ s.t. } \boldsymbol{C}^H \boldsymbol{w}_{U,L} = \boldsymbol{g}_{U,L}. \tag{49}$$

Consequently,

$$\begin{aligned}\boldsymbol{w}_{X,L}^H \boldsymbol{a} &= \xi a_L, \quad \boldsymbol{w}_{X,L}^H \boldsymbol{b} = 0,\\ \boldsymbol{w}_{X,R}^H \boldsymbol{a} &= \xi a_R, \quad \boldsymbol{w}_{X,R}^H \boldsymbol{b} = 0,\\ \boldsymbol{w}_{U,L}^H \boldsymbol{a} &= 0, \quad \boldsymbol{w}_{U,L}^H \boldsymbol{b} = \eta b_L,\\ \boldsymbol{w}_{U,R}^H \boldsymbol{a} &= 0, \quad \boldsymbol{w}_{U,R}^H \boldsymbol{b} = \eta b_R.\end{aligned} \tag{50}$$

Hence, the BLCMV beamformer is decomposed into two beamformers. The D-BLCMV beamformer lies in the null space of the constraint subspace of the interfering sources, and it can hence control the binaural cues of the desired source without affecting the interfering source component. The U-BLCMV beamformer lies in the null space of the constraint subspace of the desired sources, and it can hence control the binaural cues of the interfering source without affecting the desired source component. Based on these arguments, it is evident that the BLCMV beamformer is capable of preserving the binaural cues of both the desired and interfering sources.

F. Filter Decomposition into Two MVDR Beamformers

The left and right filters of the D-BLCMV beamformer can be written as (see Appendix A)

$$\boldsymbol{w}_{X,L} = \frac{\xi a_L^*}{1-\Gamma}\left[\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{\Gamma \boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_{ab}}\right],$$
$$\boldsymbol{w}_{X,R} = \frac{\xi a_R^*}{1-\Gamma}\left[\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{\Gamma \boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_{ab}}\right]. \quad (51)$$

Similarly, the left and right filters of the U-BLCMV beamformer can be written as

$$\boldsymbol{w}_{U,L} = \frac{\eta b_L^*}{1-\Gamma}\left[\frac{\boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_b} - \frac{\Gamma \boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_{ab}{}^*}\right],$$
$$\boldsymbol{w}_{U,R} = \frac{\eta b_R^*}{1-\Gamma}\left[\frac{\boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_b} - \frac{\Gamma \boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_{ab}{}^*}\right]. \quad (52)$$

By substituting (51) and (52) into (45) and rearranging the terms, the left and right filters of the BLCMV beamformer are given by

$$\boldsymbol{w}_L = \left[\xi a_L^* - \frac{\eta b_L^* \gamma_{ab}}{\gamma_b}\right]\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{(1-\Gamma)\gamma_a} + \left[\eta b_L^* - \frac{\xi a_L^* \gamma_{ab}{}^*}{\gamma_b}\right]\frac{\boldsymbol{R}_N^{-1}\boldsymbol{b}}{(1-\Gamma)\gamma_b},$$
$$\boldsymbol{w}_R = \left[\xi a_R^* - \frac{\eta b_R^* \gamma_{ab}}{\gamma_b}\right]\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{(1-\Gamma)\gamma_a} + \left[\eta b_R^* - \frac{\xi a_R^* \gamma_{ab}{}^*}{\gamma_b}\right]\frac{\boldsymbol{R}_N^{-1}\boldsymbol{b}}{(1-\Gamma)\gamma_b}. \quad (53)$$

Hence, it is evident that the left and right filters of the BLCMV beamformer are a linear combination of two beamformers, $\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a}$ and $\frac{\boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_b}$, which are respectively, the MVDR beamformers steered towards the desired and interfering sources.

G. Filter Decomposition Using Binauralization Postfiltering

We can further examine the D-BLCMV beamformer. Using (51), the D-BLCMV beamformer can be decomposed as

$$\boldsymbol{w}_{X,L} = d_L^* \boldsymbol{w}_X, \quad \boldsymbol{w}_{X,R} = d_R^* \boldsymbol{w}_X, \quad (54)$$

with a common filter $\boldsymbol{w}_X$ denoted as the common D-BLCMV beamformer (CD-BLCMV)

$$\boldsymbol{w}_X = \frac{1}{1-\Gamma}\boldsymbol{R}_N^{-1}\left[\frac{1}{\gamma_a}\boldsymbol{a} - \frac{\Gamma}{\gamma_{ab}}\boldsymbol{b}\right], \quad (55)$$

and $d_L = \xi a_L$, $d_R = \xi a_R$. Similarly, using (52), the U-BLCMV beamformer can be decomposed as

$$\boldsymbol{w}_{U,L} = u_L^* \boldsymbol{w}_U, \quad \boldsymbol{w}_{U,R} = u_R^* \boldsymbol{w}_U, \quad (56)$$

with a common filter $\boldsymbol{w}_U$ denoted as the common U-BLCMV beamformer (CU-BLCMV)

$$\boldsymbol{w}_U = \frac{1}{1-\Gamma}\boldsymbol{R}_N^{-1}\left[\frac{1}{\gamma_b}\boldsymbol{b} - \frac{\Gamma}{\gamma_{ab}}\boldsymbol{a}\right], \quad (57)$$

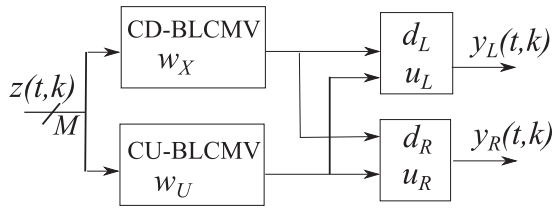

Fig. 2. The BLCMV beamformer decomposition scheme in (58).

and $u_L = \xi b_L$, $u_R = \xi b_R$. As a result, another way to describe the BLCMV beamformer is

$$\boldsymbol{w}_L = d_L^* \boldsymbol{w}_X + u_L^* \boldsymbol{w}_U, \quad \boldsymbol{w}_R = d_R^* \boldsymbol{w}_X + u_R^* \boldsymbol{w}_U. \quad (58)$$

The BLCMV beamformer can hence be decomposed into spatial filters $\boldsymbol{w}_X$ and $\boldsymbol{w}_U$, followed by single-channel postfilters d_L, d_R, u_L and u_R, respectively (see Fig. 2). Note that the ratio of d_L and d_R is equal to the input ITF of the desired source in (14), hence serving as the *desired binauralization factor*, and the multiplication of the CD-BLCMV filter with d_L^* and d_R^* as *binauralization filtering*. Similarly, the ratio of u_L and u_R is equal to the input ITF of the interfering source in (14), hence serving as the *interference binauralization factor*, and the multiplication of the CU-BLCMV filter with u_L^* and u_R^* as *binauralization filtering*.

An efficient implementation of the BLCMV beamformer can hence be obtained by sharing the common blocks given by (55) and (57). Note that the left and right filters of the D-BLCMV beamformer $\boldsymbol{w}_{X,L}$ and $\boldsymbol{w}_{X,R}$ are parallel resulting in at their output a coherent residual noise parallel to the desired source. Similarly, the left and right filters of the U-BLCMV beamformer $\boldsymbol{w}_{U,L}$ and $\boldsymbol{w}_{U,R}$ resulting in at their output a coherent residual noise parallel to the interfering source. However, the left and right filters of the BLCMV beamformer $\boldsymbol{w}_L$ and $\boldsymbol{w}_R$ which are a weighted sum of $\boldsymbol{w}_X$ and $\boldsymbol{w}_U$, are in general not parallel, and hence attributed with a non-coherent residual noise at their output.

IV. BLCMV Performance for DS Scenario

In this section, we derive analytical expressions for the performance of the BLCMV beamformer for the DS scenario in terms of binaural cue preservation, noise reduction and interference reduction. Furthermore, we provide several considerations for setting the cue gain factor for the interfering source.

A. Power Spectral Density

Substituting (13) into (11) and imposing the constraints in (21), the output PSD of the desired and the interfering source components for the left filter of the BLCMV beamformer can be computed as

$$S_{X,L,\mathrm{OUT}} = \xi^2 |a_L|^2 P_S = \xi^2 \boldsymbol{e}_L^H \boldsymbol{R}_X \boldsymbol{e}_L,$$
$$S_{U,L,\mathrm{OUT}} = \eta^2 |b_L|^2 P_U = \eta^2 \boldsymbol{e}_L^H \boldsymbol{R}_U \boldsymbol{e}_L. \quad (59)$$

The output PSD of the background noise component for the left filter of the BLCMV beamformer is derived in Appendix B, and

is given by

$$S_{N,L,\mathrm{OUT}} = \xi^2 \boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L, \tag{60}$$

with

$$\boldsymbol{R}_{XU} = \frac{1}{1-\Gamma}\left[\frac{\boldsymbol{a}\boldsymbol{a}^H}{\gamma_a} + \frac{\eta^2}{\xi^2}\frac{\boldsymbol{b}\boldsymbol{b}^H}{\gamma_b} - \Gamma\frac{\eta}{\xi}\left(\frac{\boldsymbol{a}\boldsymbol{b}^H}{\gamma_{ab^*}} + \frac{\boldsymbol{b}\boldsymbol{a}^H}{\gamma_{ab}}\right)\right], \tag{61}$$

and $\boldsymbol{R}_{XU}$ denoting the DS cross correlation matrix. Note that $\boldsymbol{R}_{XU}$ is a quadratic function in the ratio of the cue gain factors $\frac{\eta}{\xi}$. Similarly, the output PSDs of the source components in the right filter of the BLCMV beamformer are given by

$$\begin{aligned} S_{X,R,\mathrm{OUT}} &= \xi^2 |a_R|^2 P_S = \xi^2 \boldsymbol{e}_R^H \boldsymbol{R}_X \boldsymbol{e}_R, \\ S_{U,R,\mathrm{OUT}} &= \eta^2 |b_R|^2 P_U = \eta^2 \boldsymbol{e}_R^H \boldsymbol{R}_U \boldsymbol{e}_R, \\ S_{N,R,\mathrm{OUT}} &= \xi^2 \boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_R. \end{aligned} \tag{62}$$

As evident from (59), (60) and (62), the output PSDs are controlled by the correlation matrices $\boldsymbol{R}_X$, $\boldsymbol{R}_U$ and $\boldsymbol{R}_{XU}$ and the cue gain factors ξ and η.

C. Binaural Cue Preservation

Using (13) and substituting (41) and (42) into the constraint sets (21), the output ITF of the desired source for the BLCMV beamformer is equal to

$$\mathrm{ITF}_{X,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{R}_X \boldsymbol{w}_L}{\boldsymbol{w}_R^H \boldsymbol{R}_X \boldsymbol{w}_L} = \frac{\xi^2 P_S a_L a_L^*}{\xi^2 P_S a_R a_L^*} = \frac{a_L}{a_R} = \mathrm{ITF}_{X,\mathrm{IN}}, \tag{63}$$

while the output ITF of the interfering source for the BLCMV beamformer is equal to

$$\mathrm{ITF}_{U,\mathrm{OUT}} = \frac{\boldsymbol{w}_L^H \boldsymbol{R}_U \boldsymbol{w}_L}{\boldsymbol{w}_R^H \boldsymbol{R}_U \boldsymbol{w}_L} = \frac{\eta^2 P_U b_L b_L^*}{\eta^2 P_U b_R b_L^*} = \frac{b_L}{b_R} = \mathrm{ITF}_{U,\mathrm{IN}}. \tag{64}$$

Hence, the BLCMV beamformer perfectly preserves the ITF of both the desired and the interfering sources, which are both constrained[2]. The output ITF and output IC of the background noise for the BLCMV beamformer are derived in Appendix C and are given by

$$\begin{aligned} \mathrm{ITF}_{N,\mathrm{OUT}} &= \frac{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}{\boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}, \\ \mathrm{IC}_{N,\mathrm{OUT}} &= \frac{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_R}{\sqrt{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}\sqrt{\boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_R}}, \end{aligned} \tag{65}$$

and the output MSC of the background noise is equal to $|\mathrm{IC}_{N,\mathrm{OUT}}|^2$.

Many other binaural beamformers (e.g. the binaural MVDR and the binaural MWF [10]) are imposing the output ITF of all sources components (including the noise) to be equal to the input ITF of the desired source component, i.e. the noise component at the output is perceived as coming from the desired source direction. These beamformers are characterized by parallel left and right filters, hence, the output noise component is coherent and attributed with an MSC equal to one. From (61) and (65), it is evident that for the BLCMV beamformer the output ITF of the background noise depends on the input correlation matrix of the noise component and the ATFs of the desired and interfering sources. The proposed BLCMV beamformer is characterized by non-parallel filters, and therefore, the output noise component is non-coherent and attributed with an MSC smaller than one. The output noise component is characterized by a rank-2 correlation matrix, since $\boldsymbol{R}_{XU}$ is constructed from the two constrained ATFs $\boldsymbol{a}$ and $\boldsymbol{b}$.

[2]It can be shown that the BLCMV beamformer perfectly preserves the ITF of both the desired and the interfering sources for the generalized scenario with multiple sources. We omit this proof due to space constraints.

C. Interference and Noise Reduction Performance

The signal-to-interference ratio (SIR) is defined as the ratio of the PSDs of the desired and interfering source components. Using (11) and (13), the input SIR is equal to

$$\mathrm{SIR}_{L,\mathrm{IN}} = \frac{P_S |a_L|^2}{P_U |b_L|^2}, \quad \mathrm{SIR}_{R,\mathrm{IN}} = \frac{P_S |a_R|^2}{P_U |b_R|^2}, \tag{66}$$

and using (59) and (62), the output SIR is equal to

$$\mathrm{SIR}_{L,\mathrm{OUT}} = \frac{P_S |a_L|^2}{P_U |b_L|^2}\frac{\xi^2}{\eta^2}, \quad \mathrm{SIR}_{R,\mathrm{OUT}} = \frac{P_S |a_R|^2}{P_U |b_R|^2}\frac{\xi^2}{\eta^2}. \tag{67}$$

Hence, the SIR improvement is fully controlled by the cue gain factors ξ and η, i.e.

$$\Delta\mathrm{SIR}_L = \Delta\mathrm{SIR}_R = \frac{\xi^2}{\eta^2}. \tag{68}$$

The signal-to-noise ratio (SNR) is defined as the ratio of the PSDs of the desired and background noise components. Using (11) and (13), the input SNR is equal to

$$\mathrm{SNR}_{L,\mathrm{IN}} = \frac{P_S |a_L|^2}{\boldsymbol{e}_L^H \boldsymbol{R}_N \boldsymbol{e}_L}, \quad \mathrm{SNR}_{R,\mathrm{IN}} = \frac{P_S |a_R|^2}{\boldsymbol{e}_R^H \boldsymbol{R}_N \boldsymbol{e}_R}, \tag{69}$$

and using (59), (60), and (62), the output SNR is equal to

$$\mathrm{SNR}_{L,\mathrm{OUT}} = \frac{P_S |a_L|^2}{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}, \quad \mathrm{SNR}_{R,\mathrm{OUT}} = \frac{P_S |a_R|^2}{\boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_R}. \tag{70}$$

Therefore, the SNR improvement for the left and right filters of the BLCMV beamformer is given by

$$\Delta\mathrm{SNR}_L = \frac{\boldsymbol{e}_L^H \boldsymbol{R}_N \boldsymbol{e}_L}{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}, \quad \Delta\mathrm{SNR}_R = \frac{\boldsymbol{e}_R^H \boldsymbol{R}_N \boldsymbol{e}_R}{\boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_R}. \tag{71}$$

D. Setting the Cue Gain Factor for the Interfering Source

When setting the cue gain factors for the desired and the interfering sources ξ and η, different considerations need to be taken into account, e.g. based on the desired SIR and SNR improvement and the effect of RTF estimation errors. In this

subsection, we will examine the setting of η, assuming that $\xi = 1$.

First, for a scenario with a dominant interfering source, it seems desirable to set η to '0', i.e. to steer a perfect null towards the interfering source. The BLCMV beamformer then reduces to the D-BLCMV beamformer in (51) and can be rewritten as

$$\begin{aligned}\boldsymbol{w}_L &= \frac{a_L^* \boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{a_L^*\Gamma}{1-\Gamma}\left(\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{\boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_{ab}}\right),\\ \boldsymbol{w}_R &= \frac{a_R^* \boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{a_R^*\Gamma}{1-\Gamma}\left(\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{\boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_{ab}}\right).\end{aligned} \tag{72}$$

For this case, the left and right filters of the BLCMV beamformer are parallel, i.e. $\boldsymbol{w}_L = (\mathrm{ITF}_{X,\mathrm{IN}})^* \boldsymbol{w}_R$, and all signals at the output of the beamformer are perceived as arriving from the desired source direction. Hence, the ITF of all sources is equal to $\mathrm{ITF}_{X,\mathrm{IN}}$ and the MSC is equal to '1'. The output SNR for the left and right filters in (70) can now be simplified to

$$\begin{aligned}\mathrm{SNR}_{L,\mathrm{OUT}} &= \mathrm{SNR}_{R,\mathrm{OUT}} = P_S\gamma_a(1-\Gamma)\\ &= P_S\gamma_a \sin^2(\theta_{ab}),\end{aligned} \tag{73}$$

with θ_{ab} the angle between $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ in (44). While the first component in (72) is the binaural MVDR, the aim of the second component is to cancel the interfering source at the cost of reducing the SNR performance in (73) by $\Gamma P_S\gamma_a$. When $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ are orthogonal, i.e. $\theta_{ab} = \pi/2$ and, hence, $\gamma_{ab} = 0$ and $\Gamma = 0$, the BLCMV beamformer reduces to the binaural MVDR beamformer, i.e.

$$\boldsymbol{w}_L = \frac{a_L^* \boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a}, \quad \boldsymbol{w}_R = \frac{a_R^* \boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a}, \tag{74}$$

and the output SNRs of the BLCMV beamformer $\mathrm{SNR}_{L,\mathrm{OUT}}$ and $\mathrm{SNR}_{R,\mathrm{OUT}}$ in (73) are equal to $P_S\gamma_a$. When $\bar{\boldsymbol{a}}$ and $\bar{\boldsymbol{b}}$ are parallel (i.e. one is a scalar multiplier of the other), $\theta_{ab} = 0$ and $\Gamma = 1$, such that the two constraints are contradicting.

A second consideration that needs to be taken into account is the presence of RTF estimation errors, which will always arise in practice (cf. Section V). Although the D-BLCMV beamformer in (72) aims to entirely suppress the interfering source, it should be noted that when RTF estimation errors occur, the interfering source is not entirely suppressed and the residual interference leakage will be perceived as arriving from the desired source direction (cf. Section IV-B). To optimally exploit the benefits of binaural unmasking, it is desirable that the interfering source is perceived as arriving from the interfering source direction. This can be achieved by setting η to a value larger than zero where η needs to be set in accordance with the amount of residual interference leakage. On the one hand, setting η to a small value will not mask the residual interference leakage and will distort the binaural cues of the interfering source. On the other hand, setting η to a large value sacrifices SIR improvement, cf. (68). In the simulation experiments in Section VI-B, it is shown that $\eta = 0.2 - 0.3$ presents a good compromise for the considered acoustic scenario and estimation errors. In addition to residual interference leakage, it should be noted that also residual desired source leakage will occur at the output of the U-BLCMV beamformer. However, since η is typically small the residual desired source leakage can be assumed to be negligible, i.e. masked by the desired source component at the output of the D-BLCMV beamformer.

A third consideration in setting η is based on the desired SNR and SIR improvement. Substituting (61) into (70), the left output SNR, the right output SNR, and the average output SNR are given by

$$\begin{aligned}\mathrm{SNR}_{L,\mathrm{OUT}} &= \frac{P_S\gamma_a(1-\Gamma)}{\alpha_L\eta^2 - 2\beta_L\eta + 1},\\ \mathrm{SNR}_{R,\mathrm{OUT}} &= \frac{P_S\gamma_a(1-\Gamma)}{\alpha_R\eta^2 - 2\beta_R\eta + 1},\\ \mathrm{SNR}_{\mathrm{OUT}} &= \frac{S_{X,L,\mathrm{OUT}} + S_{X,R,\mathrm{OUT}}}{S_{N,L,\mathrm{OUT}} + S_{N,R,\mathrm{OUT}}} = \frac{P_S\gamma_a(1-\Gamma)}{\alpha\eta^2 - 2\beta\eta + 1},\end{aligned} \tag{75}$$

with

$$\begin{aligned}\alpha_L &= \frac{\gamma_a}{\gamma_b}\frac{|b_L|^2}{|a_L|^2}, \quad \beta_L = \Gamma Re\left\{\frac{a_L b_L^*}{{\gamma_{ab}}^*}\right\},\\ \alpha_R &= \frac{\gamma_a}{\gamma_b}\frac{|b_R|^2}{|a_R|^2}, \quad \beta_R = \Gamma Re\left\{\frac{a_R b_R^*}{{\gamma_{ab}}^*}\right\},\\ \alpha &= \frac{\gamma_a}{\gamma_b}\frac{|b_L|^2 + |b_R|^2}{|a_L|^2 + |a_R|^2}, \quad \beta = \Gamma Re\left\{\frac{a_L b_L^* + a_R {b_R}^*}{{\gamma_{ab}}^*}\right\}.\end{aligned} \tag{76}$$

where $Re\{\cdot\}$ denotes the real component. Maximizing the output SNR is obtained by setting the derivative of (75) with respect to η to zero. The cue gain factors maximizing the left output SNR, the right output SNR, and the average output SNR are given by $\eta_L = \frac{\beta_L}{\alpha_L}$, $\eta_R = \frac{\beta_R}{\alpha_R}$, and $\eta_{\mathrm{av}} = \frac{\beta}{\alpha}$, respectively. Moreover, besides SNR improvement, the SIR improvement in (68) is controlled by η. If the power of the interfering source component at the output needs to be limited, an upper limit for the value of η needs to be set.

A last consideration in setting η is the correspondence between frequency bins. The optimum η in terms of SNR is frequency-dependent. It is, however, desirable to set the same value for all frequencies in order to avoid distortion of the interfering source component at the output of the BLCMV beamformer.

V. Estimation Procedure

In the previous sections, we showed that in order to implement the BLCMV beamformer, it is sufficient to estimate the background noise correlation matrix $\boldsymbol{R}_N$ together with the subspaces that span the RTFs of the desired and interfering sources $\boldsymbol{Q}_X$ and $\boldsymbol{Q}_U$, where for the DS scenario these subspaces reduce to the individual RTFs. In this section, we describe the estimation procedure used in our implementation [31] and its limitations, together with the required assumptions regarding the activity of the sources.

For the estimation procedure three training sections are required. The first training section consists of segments in which none of the constrained sources is active, i.e. noise-only segments. These segments are used to estimate the background

noise correlation matrix $\boldsymbol{R}_N$. The second training section consists of segments in which the desired sources are active while all interfering sources are inactive, i.e. noisy desired source segments. These segments are used to estimate the noisy desired source correlation matrix, i.e. $\boldsymbol{R}_X + \boldsymbol{R}_N$. The third training section consists of segments in which the interfering sources are active while all desired sources are inactive, i.e. the noisy interference segments. These segments are used to estimate the overall noise correlation matrix, i.e. $\boldsymbol{R}_V = \boldsymbol{R}_U + \boldsymbol{R}_N$. The desired source subspace $\boldsymbol{Q}_X$ is estimated by selecting the major generalized eigenvectors of the generalized eigenvalue decomposition (GEVD) of $\boldsymbol{R}_X + \boldsymbol{R}_N$ and $\boldsymbol{R}_N$. Similarly, the interfering source subspace $\boldsymbol{Q}_U$ is estimated by selecting the major generalized eigenvectors of the GEVD of $\boldsymbol{R}_V$ and $\boldsymbol{R}_N$. In order to classify the three required training sections, we assume that an activity indicator is available, consisting of an ideal voice activity detector (VAD) and a classifier to distinguish between the three training sections and segments where desired and interfering sources are concurrently active. The implementation of such an indicator is beyond the scope of this paper, but we refer the reader to [39] for a possible solution.

The estimation procedure further relies on the following three assumptions: 1) a sufficiently large input SNR in the training sections, 2) stationarity of the background noise, and 3) static acoustic scenarios, where both the desired and interfering sources and the hearing aid user are static. In practice, RTF estimation errors are unavoidable for several reasons: Firstly, VAD errors will introduce estimation errors for the correlation matrices $\boldsymbol{R}_N$, $\boldsymbol{R}_X + \boldsymbol{R}_N$ and $\boldsymbol{R}_V$. Secondly, if the input SNR is low, it will be difficult in the GEVD procedure to distinguish between the eigenvectors belonging to the desired and interfering source subspaces, respectively, and the noise subspace. Thirdly, if the background noise is non-stationary, the background noise correlation matrix in the noisy desired and noisy interference segments, i.e. $\boldsymbol{R}_X + \boldsymbol{R}_N$ and $\boldsymbol{R}_V = \boldsymbol{R}_U + \boldsymbol{R}_N$, will be different from the background noise correlation matrix $\boldsymbol{R}_N$ estimated in the noise-only segments, leading to estimation errors in the GEVD procedure. Finally, for dynamic scenarios with moving speakers, the desired and interfering source subspaces are changing over time, such that a mechanism for tracking these subspaces would be required. In [36], such a mechanism, based on the projection approximation subspace tracking with deflation (PASTd) procedure, was proposed for a monaural version of the BLCMV beamformer. In addition, it should be noted that in [40] it was found that a monaural version of the BLCMV beamformer is quite robust to slight movements of the speakers and the hearing aid user.

RTF estimation errors will degrade the performance of the BLCMV beamformer in two aspects. Firstly, RTF estimation errors of the interfering source will lead to residual interference leakage at the output of the D-BLCMV beamformer (perceived from the direction of the desired source), while RTF estimation errors of the desired source will lead to residual desired source leakage at the output of the U-BLCMV beamformer (perceived from the direction of the interfering source), leading to a decreased SIR improvement as well as to binaural cue distortion for both sources (cf. Section VI-B). However, as already mentioned in Section IV-D, for small values of η the residual desired source leakage is masked by the desired source component at the output of the D-BLCMV beamformer. In addition, by setting η in accordance with the amount of estimation errors, it is also possible to mask the residual interference leakage by the interfering source component at the output of the U-BLCMV beamformer. Secondly, due to the constraints in (63) and (64), RTF estimation errors will lead to distorted binaural cues being imposed at the output of the BLCMV beamformer, both for the desired source as well as for the interfering source.

VI. Experimental Verification

In this section, we validate the analytical expressions for the BLCMV beamformer derived in Section IV, both using acoustic transfer functions and using sound recordings measured on Behind-the-Ear (BTE) hearing aids [41]. In Section VI-A, we examine the performance of the BLCMV beamformer in a non-reverberant environment using measured ATFs, i.e. without estimation errors. In Section VI-B, we examine the performance of the BLCMV beamformer in a reverberant environment using simulated and recorded signals, i.e. estimation errors are present.

A. Measured ATFs in a Non-Reverberant Environment

In this section, we examine the performance of the BLCMV beamformer using measured binaural Behind-the-Ear impulse responses (BTE-IRs) on two hearing aids, in a non-reverberant environment [41], at a sampling frequency of 48 kHz. Each hearing aid is equipped with 3 microphones mounted on an artificial head, i.e. all experiments were carried out using all $M = 6$ microphones. We consider two acoustic scenarios with one desired source and one interfering source for different types of background noise (either a directional noise source or diffuse noise). A general expression for the background noise correlation matrix is given by

$$\boldsymbol{R}_N = P_{N,\mathrm{DIR}}\boldsymbol{c}\boldsymbol{c}^H + P_{N,\mathrm{W}}\boldsymbol{I}_M + \boldsymbol{R}_{N,\mathrm{DIF}}, \tag{77}$$

where $\boldsymbol{c}$ is the ATF of the directional noise source, $P_{N,\mathrm{DIR}}$, and $P_{N,\mathrm{W}}$ denote the PSDs of the directional noise source, and the spatially uncorrelated white noise (e.g. sensor noise), respectively, $\boldsymbol{I}_M$ is the $M \times M$-dimensional identity matrix, and $\boldsymbol{R}_{N,\mathrm{DIF}}$ is the diffuse noise correlation matrix. The ATFs $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{c}$ of the directional sources were calculated from the BTE-IRs with an FFT window length of 2048 points. The desired source and the interfering source correlation matrices were implemented using (13). For simulating diffuse noise, a cylindrically isotropic noise field was assumed. The (i,j)-th element of the noise correlation matrix $\boldsymbol{R}_{N,\mathrm{DIF}}^{i,j}$ was calculated using the ATFs of the anechoic BTE-IRs as

$$\boldsymbol{R}_{N,\mathrm{DIF}}^{i,j} = P_{N,\mathrm{DIF}}\frac{\sum_{n=1}^{N} H_i(\theta_n)H_j^*(\theta_n)}{\sqrt{\sum_{n=1}^{N}|H_i(\theta_n)|^2\sum_{n=1}^{N}|H_j(\theta_n)|^2}}, \tag{78}$$

with $P_{N,\mathrm{DIF}}$ denote the PSD of the diffuse noise, $H(\theta_n)$ denoting the measured ATF at angle θ_n and N the total number of

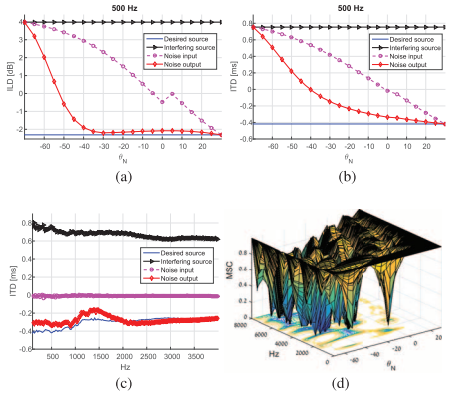

Fig. 3. Binaural cues and MSC for a directional noise source as a function of noise source angle θ_N for 500 Hz (a)-(c) and (d) ITD as a function of frequency for $\theta_N = 0^\circ$ (desired source at 30°, interfering source at -70°, $\xi = 1$, $\eta = 0.2$, $M = 6$). (a) ILD (b) ITD (c) ITD (d) MSC.

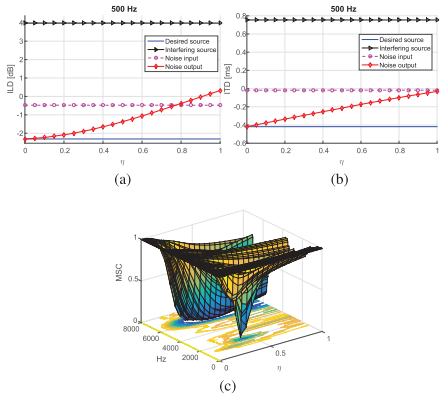

Fig. 4. Binaural cues and MSC for a directional noise source at $\theta_N = 0^\circ$ as a function of η (desired source at 30°, interfering source at -70°, $\xi = 1$, $M = 6$). (a) ILD (b) ITD (c) MSC.

angles ($N = 72$). The MSC of the diffuse noise at the reference microphones is depicted in Fig. 5(c). To verify the theoretical analysis presented in Section IV, we circumvented any estimation error issues.

The desired and interfering sources were located at angles $\theta_X = 30^\circ$ and $\theta_U = -70^\circ$ from the artificial head, respectively (angle $\theta = 0^\circ$ denotes a signal arriving from the front, and angle $\theta = 90^\circ$ from the right). Note that sensor noise, modeled as spatially uncorrelated white noise, was added to $\boldsymbol{R}_N$, i.e. $P_{N,\mathrm{W}}$ is set 55 dB lower than the desired source PSD. This also ensures that the noise correlation matrix, $\boldsymbol{R}_N$, is invertible. The ratio of the PSD of the desired source to the PSD of the interfering source was set to 0 dB. We set $\xi = 1$ for all scenarios, while η varies between 0 to 1.

Without estimation errors, the binaural cues of the desired and interfering sources are perfectly preserved by the BLCMV beamformer, as was derived in (63) and (64). Hence, for the sake of brevity, only the binaural cues of the unconstrained noise (either directional noise source or diffuse noise) at the output of the beamformer are examined.

1) Directional Noise Source: In the first scenario, we consider two dominant directional undesired sources. The directional interfering source is constrained by the BLCMV beamformer, while the directional noise source is unconstrained. For $\eta = 0.2$, Fig. 3(a) and Fig. 3(b) depict $\mathrm{ILD}_{N,\mathrm{OUT}}$ and $\mathrm{ITD}_{N,\mathrm{OUT}}$, i.e. the binaural cues of the noise source at the output of the beamformer, at 500 Hz, as a function of θ_N for $\theta_U \leq \theta_N \leq \theta_X$ (i.e. the direction of the directional noise source changes between the direction of the desired source and the interfering source). It can be observed that the binaural cues of the (unconstrained) noise source are distorted and vary as a function of its direction. Although, not shown in the figure, we note that the binaural cues of the (constrained) interfering source are preserved, cf. (64).

Fig. 3(c) depicts the frequency-dependent ITDs for the desired source, the interfering source and the noise source for $\theta_N = 0^\circ$ and $\eta = 0.2$. Again, it can be clearly observed that the ITD of the noise source is not preserved. Similar conclusions can be drawn for the frequency-dependent ILD of the noise source. Fig. 3(d) depicts $\mathrm{MSC}_{N,\mathrm{OUT}}$, i.e. the MSC of the noise source at the output of the beamformer. The MSC of the noise is lower than one and as a result the noise is perceived as non-coherent. It is evident that when the direction of the noise source is close to the direction of the desired source (i.e. θ_N is close to 30°), $\mathrm{MSC}_{N,\mathrm{OUT}}$ is approximately equal to one. At low frequencies $\mathrm{MSC}_{N,\mathrm{OUT}}$ varies moderately, whereas at high frequencies $\mathrm{MSC}_{N,\mathrm{OUT}}$ varies rapidly as a function of θ_N.

For a directional noise source at $\theta_N = 0^\circ$ we further examine the binaural cues $\mathrm{ILD}_{N,\mathrm{OUT}}$, $\mathrm{ITD}_{N,\mathrm{OUT}}$ and $\mathrm{MSC}_{N,\mathrm{OUT}}$ for different values of η. The results are depicted in Fig. 4. As expected (cf. (72)), setting η to zero, $\mathrm{ILD}_{N,\mathrm{OUT}} = \mathrm{ILD}_X$, $\mathrm{ITD}_{N,\mathrm{OUT}} = \mathrm{ITD}_X$ and $\mathrm{MSC}_{N,\mathrm{OUT}} = 1$. However, as η increases, $\mathrm{ILD}_{N,\mathrm{OUT}}$ and $\mathrm{ITD}_{N,\mathrm{OUT}}$ are shifted towards ILD_U and ITD_U, respectively, and $\mathrm{MSC}_{N,\mathrm{OUT}}$ is lower than one, varying for different frequencies.

2) Diffuse Noise: In the second scenario, we consider one dominant directional undesired source (i.e. the interfering source) in a diffuse noise environment. This scenario is encountered, for example, in a car.

Fig. 5(a), Fig. 5(b), and Fig. 5(c) depict $\mathrm{ILD}_{N,\mathrm{OUT}}$, $\mathrm{ITD}_{N,\mathrm{OUT}}$ and $\mathrm{MSC}_{N,\mathrm{OUT}}$ as a function of frequency, for $\eta = 0.2$. Fig. 5(a) and Fig. 5(b) show that $\mathrm{ILD}_{N,\mathrm{OUT}}$ and $\mathrm{ITD}_{N,\mathrm{OUT}}$ are approximately equal to $\mathrm{ILD}_{X,\mathrm{IN}}$ and $\mathrm{ITD}_{X,\mathrm{IN}}$, respectively. From Fig. 5(c), it can be observed that although the noise is diffuse at the input of the beamformer, the MSC is quite different at the output of the beamformer.

The binaural cues of the diffuse noise were further examined for different values of η. The $\mathrm{ILD}_{N,\mathrm{OUT}}$ and $\mathrm{ITD}_{N,\mathrm{OUT}}$ results are similar to the directional noise source results, and hence are not shown. Fig. 5(d) depicts $\mathrm{MSC}_{N,\mathrm{OUT}}$ as a function of η and frequency. Note that Fig. 5(c) is a snapshot of Fig. 5(d) for $\eta = 0.2$. It is evident that for $\eta = 0$, $\mathrm{MSC}_{N,\mathrm{OUT}} = 1$, as for the directional noise scenario. As η increases, $\mathrm{MSC}_{N,\mathrm{OUT}}$ decreases. At low frequencies $\mathrm{MSC}_{N,\mathrm{OUT}}$ is slightly lower than '1', whereas at higher frequencies $\mathrm{MSC}_{N,\mathrm{OUT}}$ is relatively low, implying that the noise at the output exhibits non-coherent characteristics. Since $\mathrm{MSC}_{N,\mathrm{OUT}}$ is lower than '1', the noise at the output of the beamformer is not perceived as a directional source. The fact that the residual noise at the output is non-coherent is advantageous in terms of speech intelligibility due

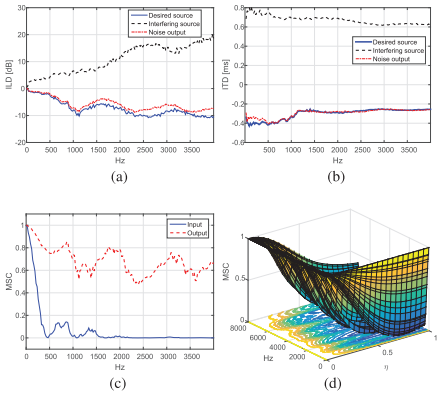

Fig. 5. Binaural cues and MSC for diffuse noise as a function of frequency (desired source at $30°$, interfering source at $-70°$, $\xi = 1$, $M = 6$). (a)-(c) with $\eta = 0.2$ and (d) MSC for various η values. (a) ILD (b) ITD (c) MSC (d) MSC.

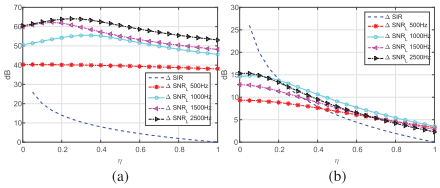

Fig. 6. SIR and SNR improvements for a directional noise source at $0°$ and diffuse noise at the left filter as a function of η for several frequencies (desired source at $30°$, interfering source at $-70°$, $\xi = 1$, $M = 6$). (a) Directional noise source (b) Diffuse noise.

to binaural unmasking, in contrast to other widely-used binaural beamformers, e.g. the binaural MVDR beamformer and the binaural MWF [10], for which $\mathrm{MSC}_{N,\mathrm{OUT}} = 1$ such that the residual noise at the output is coherent and perceived from the direction of the desired source.

3) Interference and Noise Reduction Performance: Fig. 6 depicts the narrowband SIR and SNR improvement for the left BLCMV filter, i.e. $\Delta\mathrm{SIR}_L$ and $\Delta\mathrm{SNR}_L$, as a function of η for several frequencies. Note that similar results were obtained for the right filter, where the SIR improvement for the left and right filters is equal, i.e. $\Delta\mathrm{SIR} = \Delta\mathrm{SIR}_L = \Delta\mathrm{SIR}_R$. From these figures it is evident that $\Delta\mathrm{SIR}$ is inversely proportional to η. Furthermore, it is clear that η influences the achievable SNR improvement and that the optimum η in terms of SNR improvement is frequency-dependent. For the directional noise source (Fig. 6(a)) it can be observed that the SNR improvement is very high. When comparing the SNR improvement for the directional noise source in Fig. 6(a) to the SNR improvement for the diffuse noise in Fig. 6(b), it is evident, as expected, that while the SIR improvement for the directional noise source is similar to the SIR improvement for the diffuse noise, the SNR improvement is significantly reduced for the diffuse noise.

B. Reverberant Environment

In this section, we examine the performance of the BLCMV beamformer in a reverberant environment using simulated and recorded signals, i.e. when estimation errors are present. Similar to Section VI-A, the experimental setup consists of two hearing aid devices, each with three microphones mounted on an artificial head in a cafeteria with a reverberation time of approximately 1.25 s [41]. The directional sources are synthesized by convolving clean speech and noise signals with measured BTE-IRs for different positions in the cafeteria. Babble noise, originating from multiple simultaneous conversations constitutes the diffuse sound field.

We will consider two acoustic scenarios, either with a directional noise source or diffuse noise (cf. Table I). The first acoustic scenario, denoted S1 to S3, is comprised of one desired speaker, one interfering speaker, and one directional background noise source at various positions. The second acoustic scenario, denoted S4 and S5, is comprised of one desired speaker, one interfering speaker, and diffuse babble noise recorded in the cafeteria. The input SIR, with respect to the interfering source, and the input SNR, with respect to the background noise, were set to 6 dB and 14 dB, respectively. Additional uncorrelated white sensor noise was added to guarantee that the noise correlation matrix is always invertible. The signal-to-sensor noise ratio was set to 45 dB. The sampling frequency was 8 kHz. The signals were transformed to the STFT domain with 4096 samples per frame and 75% overlap.

The BLCMV beamformer was implemented in a generalized sidelobe canceler (GSC) form, which is equivalent to the BLCMV beamformer using $\boldsymbol{R}_N$. A detailed description of the GSC implementation can be found in [24], [31]. The RTFs of the desired and interfering source were estimated using the GEVD procedure, as described in Section V. The cue gain factor for the desired source ξ was set to one, and the cue gain factor for the interfering source η was varied between 0.2 and 0.3.

For evaluating the performance of the BLCMV beamformer, we applied the algorithm in two phases. In the first phase, the BLCMV beamformer was applied to the actual input signals, comprised of the sum of the desired source, the interfering source, and the background noise. In this phase, the beamformer was allowed to adapt yielding the actual algorithm output. In order to examine the contribution of the D-BLCMV beamformer in (51) and the U-BLCMV beamformer in (52), each decomposed beamformer was employed as well. In the second phase, the beamformers were not allowed to adapt. Instead, a copy of the time-varying filter coefficients obtained in the first phase was used. Each beamformer was applied to the desired source component, the interfering source component and the background noise component separately. This procedure enables a careful examination of the performance measures, i.e. the SIR improvement, SNR improvement and the binaural cue preservation capabilities of the beamformer. The distortion is assessed by calculating the log spectral distortion (LSD) measure relating the desired source component at the left output of the BLCMV beamformer to the left microphone reference.

Table I summarizes the performance for the various acoustic scenarios, where (θ, d) is a descriptor for a directional source at angle θ and distance d from the artificial head. Scenario S1 consists of one desired speaker at position A ($0°$, 102 cm), one interfering speaker at position D ($-90°$, 162 cm), and one directional stationary pink noise source at position E ($-135°$, 129 cm). Fig. 7 depicts the sonograms at the right hearing aid as well as both the left and right waveforms for scenario S1 for the desired source component, the interfering source component,

TABLE I
SNR AND SIR IMPROVEMENTS IN dB (RELATIVE TO THE LEFT AND RIGHT REFERENCE MICROPHONE SIGNALS) AND LSD FOR VARIOUS ACOUSTIC SCENARIOS ($T_{60} = 1250$ ms, $\xi = 1$, $M = 6$). 0° DEFINES THE POSITION IN FRONT OF THE LISTENER. THE AZIMUTH ANGLE IS DEFINED AS COUNTER-CLOCKWISE

	Desired	Interference	Noise	η	ΔSIR_L	ΔSIR_R	ΔSNR_L	ΔSNR_R	LSD
S1	A (0°, 102cm)	D (−90°, 162cm)	E (−135°, 129cm)	0.2	11.41	12.30	18.52	21.10	1.60
S2	A (0°, 102cm)	B (45°, 118cm)	D (−90°, 162cm)	0.2	10.72	9.88	13.98	17.70	1.89
S3	A (0°, 102cm)	B (45°, 118cm)	D (−90°, 162cm)	0.3	9.15	8.39	14.59	17.92	1.83
S4	A (0°, 102cm)	B (45°, 118cm)	Babble	0.3	9.78	9.23	5.84	4.95	1.56
S5	A (0°, 102cm)	B (45°, 118cm)	Babble	0	15.0	12.9	5.1	4.31	1.75

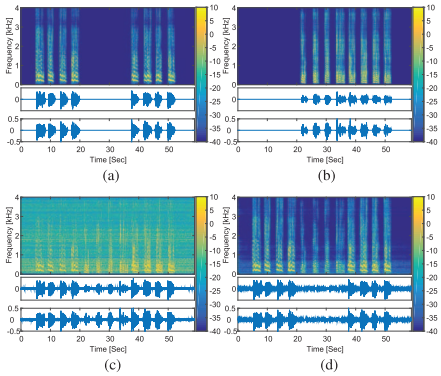

Fig. 7. Sonograms of the right signal and stereo waveforms (left and right signals) for the BLCMV beamformer for scenario S1. (a) Desired source signal (b) Interfering source signal (c) Noisy source signal (d) Enhanced source signal.

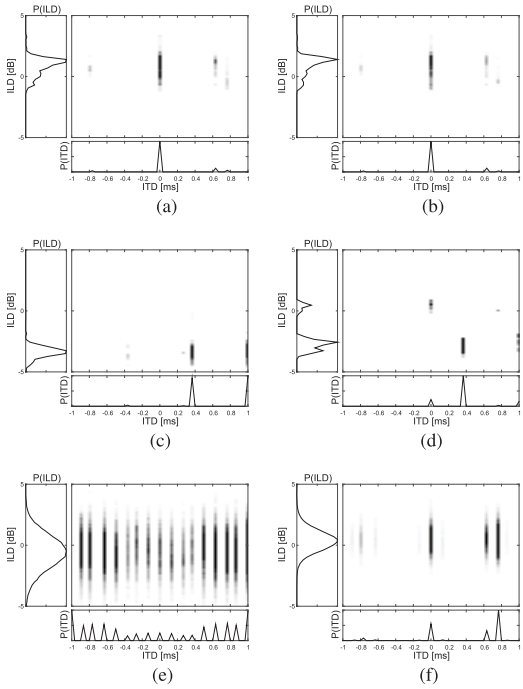

Fig. 8. The PDFs of ITD and ILD ($T_{60} = 1250$ ms, scenario S4, desired source at A (0°, 102 cm), interfering source at B (45°, 118 cm), and diffuse babble noise, $\xi = 1$, $\eta = 0.3$, $M = 6$). Data are shown for the critical band at 1480 Hz. Graphs produced by the binaural cue selection Matlab toolbox [26]. (a) Desired source at the input (b) Desired source at the output (c) Interfering source at the input (d) Interfering source at the output (e) Babble noise at the input (f) Babble noise at the output.

the noisy reference microphone signals and the BLCMV beamformer output signals. It can be observed that the BLCMV beamformer significantly attenuates the interfering source and the stationary noise. In scenario S2, we examine the performance of the BLCMV beamformer when the interfering source is close to the desired source. This scenario is comprised of one desired speaker at position A, one interfering speaker at position B (45°, 118 cm), and one directional stationary pink noise source at position D. As expected, it can be observed that the performance for scenario S1 outperforms the performance for scenario S2 in terms of SIR improvement and SNR improvement. In scenarios S1 and S2 η is equal to 0.2, corresponding to 14 dB of desired attenuation for the interfering source. Note however that due to estimation errors the actual SIR improvements are smaller than 14 dB. In scenario S3, we change η to 0.3, corresponding to 10 dB of desired attenuation for the interfering source. Note that this hardly degraded the SIR improvement compared to scenario S2. In scenario S4, diffuse babble noise is used instead of the directional noise source. Since diffuse noise can be modeled as a superposition of uncorrelated plane waves from various directions, the spatial filtering capabilities of the beamformer are expected to be limited. While the SIR improvement performance for scenario S4 is comparable to the SIR improvement for scenarios S2-S3, the SNR improvement significantly decreases. In scenario S5 we set $\eta = 0$, such that the BLCMV beamformer reduces to the D-BLCMV beamformer in (51). While the SIR improvement for scenario S5 increases compared to scenario S4, the SNR improvement is comparable. For all scenarios it can be observed that the LSD measure is comparable (lower values indicate less distortion).

The analysis of the binaural cue preservation is carried out using a modeling framework, motivated by human auditory-based processing as described in [26]. The model is based on the histogram of ITD and ILD values of time segments that have passed a predefined threshold. The threshold is set to imitate the human spatial perception of coherent sources and is based on the IC. We evaluate the binaural cue preservation of the proposed beamformers for each source separately in one critical band, centered at 1480 Hz. Fig. 8 and Fig. 9 depict the probability density functions (PDFs) of the selected ITD and ILD cues for scenario S4. The IC threshold was set to 0.993. From Fig. 8(a) and Fig. 8(b), it can be observed that the desired source component at the output of the BLCMV beamformer is preserved. Similar conclusions regarding the interfering source can be drawn from Fig. 8(c) and Fig. 8(d). However, it can be observed that the interfering source component at the output of the BLCMV beamformer includes a component with binaural cues of the desired source, which can probably attributed to a residual interference leakage due to estimation errors. In Fig. 8(e) and Fig. 8(f), the binaural cues of the diffuse noise are examined. The IC threshold was set to 0.5 (since the signal is non-coherent). It is evident that the binaural cues of the

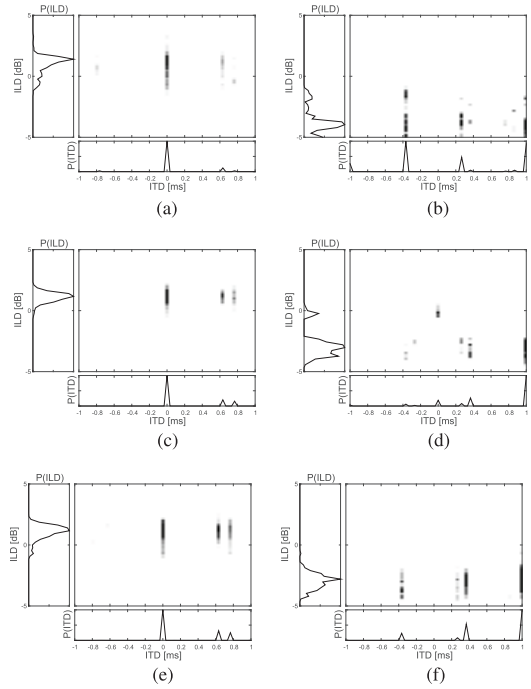

Fig. 9. The PDFs of ITD and ILD for the D-BLCMV and U-BLCMV beamformers ($T_{60} = 1250$ ms, scenario S4, desired source at A (0°, 102 cm), interfering source at B (45°, 118 cm), and diffuse babble noise, $\xi = 1$, $\eta = 0.3$, $M = 6$). Data are shown for the critical band at 1480 Hz. Graphs produced by the binaural cue selection Matlab toolbox [26]. (a) Desired source at the D-BLCMV output (b) Desired source at the U-BLCMV output (c) Interfering source at the D-BLCMV output (d) Interfering source at the U-BLCMV output (e) Noise source at the D-BLCMV output (f) Noise source at the U-BLCMV output.

(unconstrained) noise are not preserved, but rather replaced by the binaural cues of a mix of the desired and interfering sources.

In order to emphasize the contribution of the beamformer decomposition in (45), Fig. 9 depicts the binaural cues of each source separately at the output of the D-BLCMV beamformer in (51) and the U-BLCMV beamformer in (52). On the one hand, it can be observed that the binaural cues of all sources (i.e. the desired, interfering and noise sources) at the D-BLCMV beamformer output are similar to the binaural cues of the desired source at the input. On the other hand, it can be observed that the binaural cues of all sources at the output of the U-BLCMV beamformer are similar to the binaural cues of the interfering source at the input. This means that the D-BLCMV beamformer imposes the binaural cues of the desired source at the output, while the U-BLCMV beamformer imposes the binaural cues of the interfering source at the output. Particularly, it can be observed from Fig. 9(c) that the binaural cues of the residual interference leakage at the output of the D-BLCMV beamformer resemble the binaural cues of the desired source (cf. Fig. 8(a)), and are hence distorted. On the contrary, it can be observed from Fig. 9(d) that the binaural cues of the interfering source component at the output of the U-BLCMV beamformer resemble the correct binaural cues of the interfering source (cf. Fig. 8(c)). The interfering source component at the output of the BLCMV beamformer is equal to the sum of the residual interference leakage (from the D-BLCMV beamformer) and an attenuated interfering source component (from the U-BLCMV beamformer). This can be observed in Fig. 8(d), where the binaural cues of the interfering source component at the output of the BLCMV beamformer consist of a component with the binaural cues of the interfering source as well as a residual interference leakage component with the binaural cues of the desired source. However, by setting $\eta = 0.3$ (in accordance with the amount of residual interference leakage), the residual interference leakage component is perceptually masked by the interfering source component at the output of the U-BLCMV beamformer, which has been verified using informal listening tests[3].

Performance may degrade in highly reverberant environments when the frame length of the STFT window decreases. When the reverberation level increases, the relative impulse responses may become too long to be adequately modeled by the used frame length. If the relative impulse response, and hence, the corresponding RTF, is longer than the STFT window, the convolutive transfer function (CTF) approximation can be used instead [42].

VII. Conclusions

In this paper, the BLCMV beamformer was discussed. To fully exploit the advantage of binaural hearing, the BLCMV beamformer preserves the binaural cues of the constrained sources in addition to providing the undistorted extraction of the desired source and noise reduction. A theoretical analysis of the BLCMV beamformer was introduced and several filter decompositions were derived. Analytical expressions for the BLCMV performance were evaluated in terms of noise reduction, interference reduction, and cue preservation. Various considerations are taken into account when setting the BLCMV beamformer parameters, which allow to control its performance. Comprehensive simulation and experimental verifications using both measured acoustic transfer functions and real recordings exemplify the BLCMV beamformer capabilities in various noise environments.

Appendix A
D-BLCMV Filter Decomposition

In this appendix, the left and right filters of the D-BLCMV beamformer are derived. Substituting (41) and (47) into (46), the solution for the left D-BLCMV problem in (48) is given by

$$\begin{aligned}\boldsymbol{w}_{X,L} &= \left[\boldsymbol{R}_N^{-1}\boldsymbol{a}\ \boldsymbol{R}_N^{-1}\boldsymbol{b}\right]\left[\boldsymbol{C}^H\boldsymbol{R}_N^{-1}\boldsymbol{C}\right]^{-1}\begin{bmatrix}\xi a_L^*\\0\end{bmatrix}\\ &= \left[\boldsymbol{R}_N^{-1}\boldsymbol{a}\ \boldsymbol{R}_N^{-1}\boldsymbol{b}\right]\begin{bmatrix}\gamma_a & \gamma_{ab}\\ \gamma_{ab}{}^* & \gamma_b\end{bmatrix}^{-1}\begin{bmatrix}\xi a_L^*\\0\end{bmatrix}\\ &= \left[\boldsymbol{R}_N^{-1}\boldsymbol{a}\ \boldsymbol{R}_N^{-1}\boldsymbol{b}\right]\frac{1}{\gamma_a\gamma_b(1-\Gamma)}\begin{bmatrix}\gamma_b & -\gamma_{ab}\\ -\gamma_{ab}{}^* & \gamma_a\end{bmatrix}\begin{bmatrix}\xi a_L^*\\0\end{bmatrix}\\ &= \left[\boldsymbol{R}_N^{-1}\boldsymbol{a}\ \boldsymbol{R}_N^{-1}\boldsymbol{b}\right]\frac{\xi a_L^*}{\gamma_a\gamma_b(1-\Gamma)}\begin{bmatrix}\gamma_b\\ -\gamma_{ab}{}^*\end{bmatrix}. \end{aligned} \tag{79}$$

Hence, the left filter of the D-BLCMV beamformer is equal to

$$\boldsymbol{w}_{X,L} = \frac{\xi a_L^*}{1-\Gamma}\left[\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{\Gamma\boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_{ab}}\right], \tag{80}$$

[3]http://www.eng.biu.ac.il/gannot/speech-enhancement/binaural-lcmv/

and similarly, the right filter of the D-BLCMV beamformer is equal to

$$\boldsymbol{w}_{X,R} = \frac{\xi a_R^*}{1-\Gamma}\left[\frac{\boldsymbol{R}_N^{-1}\boldsymbol{a}}{\gamma_a} - \frac{\Gamma \boldsymbol{R}_N^{-1}\boldsymbol{b}}{\gamma_{ab}}\right], \tag{81}$$

where the right filter is evaluated by substituting a_L with a_R.

Appendix B
Output Noise PSD

Using (45), the output PSD of the noise component for the left filter of the BLCMV beamformer is given by

$$\begin{aligned}\boldsymbol{w}_L^H \boldsymbol{R}_N \boldsymbol{w}_L &= (\boldsymbol{w}_{X,L} + \boldsymbol{w}_{U,L})^H \boldsymbol{R}_N (\boldsymbol{w}_{X,L} + \boldsymbol{w}_{U,L}) \\ &= \xi^2 a_L \boldsymbol{w}_X^H \boldsymbol{R}_N \boldsymbol{w}_X a_L^* + \eta^2 b_L \boldsymbol{w}_U^H \boldsymbol{R}_N \boldsymbol{w}_U b_L^* \\ &\quad + \xi\eta a_L \boldsymbol{w}_X^H \boldsymbol{R}_N \boldsymbol{w}_U b_L{}^* + \xi\eta b_L \boldsymbol{w}_U{}^H \boldsymbol{R}_N \boldsymbol{w}_X a_L{}^*.\end{aligned} \tag{82}$$

Using (55) and (57), the four components in (82) are given by

$$\begin{aligned}&\boldsymbol{w}_X^H \boldsymbol{R}_N \boldsymbol{w}_X = \frac{1}{(1-\Gamma)\gamma_a}, \boldsymbol{w}_U^H \boldsymbol{R}_N \boldsymbol{w}_U = \frac{1}{(1-\Gamma)\gamma_b}, \\ &\boldsymbol{w}_X^H \boldsymbol{R}_N \boldsymbol{w}_U = -\frac{\Gamma}{(1-\Gamma)\gamma_{ab}{}^*}, \boldsymbol{w}_U^H \boldsymbol{R}_N \boldsymbol{w}_X = -\frac{\Gamma}{(1-\Gamma)\gamma_{ab}}.\end{aligned} \tag{83}$$

Substituting (83) into (82) yields

$$\begin{aligned}&\boldsymbol{w}_L^H \boldsymbol{R}_N \boldsymbol{w}_L \\ &= \frac{1}{1-\Gamma}\left[\frac{|a_L|^2\xi^2}{\gamma_a} + \frac{|b_L|^2\eta^2}{\gamma_b} - \xi\eta\Gamma\left(\frac{a_L b_L{}^*}{\gamma_{ab}{}^*} + \frac{a_L^* b_L}{\gamma_{ab}}\right)\right] \\ &= \frac{\xi^2}{1-\Gamma}\boldsymbol{e}_L^H\left[\frac{\boldsymbol{a}\boldsymbol{a}^H}{\gamma_a} + \frac{\eta^2}{\xi^2}\frac{\boldsymbol{b}\boldsymbol{b}^H}{\gamma_b} - \frac{\eta}{\xi}\Gamma\left(\frac{\boldsymbol{a}\boldsymbol{b}^H}{\gamma_{ab}{}^*} + \frac{\boldsymbol{b}\boldsymbol{a}^H}{\gamma_{ab}}\right)\right]\boldsymbol{e}_L \\ &= \xi^2 \boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L,\end{aligned} \tag{84}$$

where $\boldsymbol{R}_{XU}$ is defined in (61).

Appendix C
Noise ITF and IC

By substituting the BLCMV decomposition in (45) into (6), the ITF of the output noise component is equal to

$$\mathrm{ITF}_{N,\mathrm{OUT}} = \frac{(\boldsymbol{w}_{X,L} + \boldsymbol{w}_{U,L})^H \boldsymbol{R}_N (\boldsymbol{w}_{X,L} + \boldsymbol{w}_{U,L})}{(\boldsymbol{w}_{X,R} + \boldsymbol{w}_{U,R})^H \boldsymbol{R}_N (\boldsymbol{w}_{X,L} + \boldsymbol{w}_{U,L})}. \tag{85}$$

Similarly to the derivation in (84), using (83) yields

$$\boldsymbol{w}_R^H \boldsymbol{R}_N \boldsymbol{w}_L = \xi^2 \boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_L, \quad \boldsymbol{w}_R^H \boldsymbol{R}_N \boldsymbol{w}_R = \xi^2 \boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_R. \tag{86}$$

Hence, the noise ITF at the output of the BLCMV beamformer is equal to

$$\mathrm{ITF}_{N,\mathrm{OUT}} = \frac{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}{\boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}, \tag{87}$$

where $\boldsymbol{R}_{XU}$ is defined in (61).

The IC of the output noise component can be computed similarly. Substituting (84) and (86) into (10) yields

$$\mathrm{IC}_{N,\mathrm{OUT}} = \frac{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_R}{\sqrt{\boldsymbol{e}_L^H \boldsymbol{R}_{XU} \boldsymbol{e}_L}\sqrt{\boldsymbol{e}_R^H \boldsymbol{R}_{XU} \boldsymbol{e}_R}}. \tag{88}$$

References

[1] V. Algazi and R. Duda, "Headphone-based spatial sound," *IEEE Signal Proc. Magazine*, vol. 28, no. 1, pp. 33–42, 2011.

[2] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Acoust. Soc. Amer.*, vol. 55, no. 6, pp. 503–516, 2007.

[3] B. Kollmeier, J. Peissig, and V. Hohmann, "Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain," *Scandinavian Audiology. Supplementum*, vol. 38, p. 28, 1993.

[4] T. Wittkop and V. Hohmann, "Strategy-selective noise reduction for binaural digital hearing aids," *Speech Communication*, vol. 39, no. 1, pp. 111–138, 2003.

[5] J. Li, M. Akagi, and Y. Suzuki, "Extension of the two-microphone noise reduction method for binaural hearing aids," *Inter. Conf. on Audio, Language and Image Processing (ICALIP)*, 2008, pp. 97–101.

[6] D. Wang and G. Brown, *Computational auditory scene analysis: Principles, algorithms, and applications*, IEEE Press, 2006.

[7] S. Wehr, M. Zourub, R. Aichner, and W. Kellermann, "Post-processing for BSS algorithms to recover spatial cues," *Proc. Int. Workshop Acoust. Signal Enhance. (IWAENC)*, Paris, France, Sep. 2006.

[8] R. Aichner, H. Buchner, M. Zourub, and W. Kellermann, "Multi-channel source separation preserving spatial information," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, USA, Apr. 2007, pp. 5–8.

[9] K. Reindl, Y. Zheng, and W. Kellermann, "Speech enhancement for binaural hearing aids based on blind source separation," *4th Int. Symp. on Communications, Control and Signal Proc. (ISCCSP)*, Mar. 2010, pp. 1–6.

[10] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," *Handbook on Array Processing and Sensor Networks*, 2008, pp. 269–302.

[11] S. Doclo, R. Dong, T. Klasen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel wiener filter with itd cues for noise reduction in binaural hearing aids," *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, 2005, pp. 70–73.

[12] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[13] D. Marquardt, V. Hohmann, and S. Doclo, "Binaural cue preservation for hearing aids using multi-channel Wiener filter with instantaneous ITF preservation," *Proc. ICASSP*, Kyoto, Japan, Mar. 2012, pp. 21–24.

[14] D. Marquardt, V. Hohmann, and S. Doclo, "Coherence preservation in multichannel Wiener filtering based noise reduction for binaural hearing aids," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, Canada, May 2013, pp. 8648–8652.

[15] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Binaural multichannel Wiener filter with directional interference rejection," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 644–648.

[16] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Theoretical analysis of linearly constrained multi-channel Wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 23, no. 12, pp. 2384–2397, Dec. 2015.

[17] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Extensions of the binaural MWF with interference reduction preserving the binaural cues of the interfering source," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016.

[18] J. Desloge, W. Rabinowitz, and P. Zurek, "Microphone-array hearing aids with binaural output. I. Fixed-processing systems," *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 6, pp. 529–542, Nov. 1997.

[19] D. Welker, J. Greenberg, J. Desloge, and P. Zurek, "Microphone-array hearing aids with binaural output. II. a two-microphone adaptive system," *IEEE Trans. Speech and Audio Proc.*, vol. 5, no. 6, pp. 543–551, 1997.

[20] Y. Suzuki, S. Tsukui, F. Asano, and R. Nishimura, "New design method of a binaural microphone array using multiple constraints," *IEICE Tran. on Fundamentals of Elect., Comm. and Comp. Sci.*, vol. 82, no. 4, pp. 588–596, 1999.

[21] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Appl. Signal Process.*, pp. 175–175, Jan. 2006.

[22] J. D. Gordy, M. Bouchard, and T. Aboulnasr, "Beamformer performance limits in monaural and binaural hearing aid applications," *Canadian Conf. on Elect. and Comp. Eng. (CCECE)*, 2008, pp. 381–386.

[23] S. Markovich-Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer," *Proc. Int. Workshop Acoust. Signal Enhance. (IWAENC)*, Tel-Aviv, Israel, Sep. 2010.

[24] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," *Proc. Int. Workshop Acoust. Signal Enhance. (IWAENC)*, Aachen, Germany, Sep. 2012, pp. 117–120.
[25] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.
[26] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Amer.*, vol. 116, no. 5, pp. 3075–3089, 2004.
[27] M. Dietz, S. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Communication*, vol. 53, no. 5, pp. 592–605, 2011.
[28] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. 5, no. 2, pp. 4–24, 1988.
[29] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
[30] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech and Audio Proc.*, vol. 12, no. 5, pp. 451–459, 2004.
[31] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.
[32] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, John Wiley & Sons, 2004.
[33] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. Acoust. Soc. Amer.*, vol. 54, no. 3, pp. 771–785, 1973.
[34] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, "Sensitivity analysis of mvdr and mpdr beamformers," *26th Conv. Electr. Electron. Eng. in Israel (IEEEI)*, 2010, pp. 416–420.
[35] G. Reuven, S. Gannot, and I. Cohen, "Performance analysis of the dual source transfer-function generalized sidelobe canceler," *Speech Communication*, vol. 49, no. 7–8, pp. 602–622, Jul. 2007.
[36] S. Markovich-Golan, S. Gannot, and I. Cohen, "Subspace tracking of multiple sources and its application to speakers extraction," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2010, pp. 201–204.
[37] J. Benesty, J. Chen, and Y. Huang, "Estimation of the coherence function with the MVDR approach," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, pp. 500–503.
[38] C. Zheng, M. Zhou, and X. Li, "On the relationship of non-parametric methods for coherence function estimation," *Signal Processing*, vol. 88, no. 11, pp. 2863–2867, 2008.
[39] M. Taseska, S. Markovich-Golan, E. Habets, and S. Gannot, "Near-field source extraction using speech presence probabilities for ad hoc microphone arrays," *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Antibes - Juan les Pins, France, Sep. 2014, pp. 169–173.
[40] W. Woods, E. Hadad, I. Merks, B. Xu, S. Gannot, and T. Zhang, "A real-world recording database for ad hoc microphone arrays," *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz NY, USA, Oct. 2015, pp. 1–5.
[41] H. Kayser, S. Ewert, J. Annemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. Adv. Signal Process.*, 10, 2009.
[42] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Proc.*, vol. 17, no. 4, pp. 546–555, 2009.

Elior Hadad (S'13) received the B.Sc. (summa cum laude) and the M.Sc. (cum laude) degrees in Electrical Engineering from Ben-Gurion University of the Negev (BGU), Beer-Sheva, Israel, in 2001 and 2007, respectively. He is currently pursuing his Ph.D. degree at the Engineering Faculty, Bar-Ilan University, Israel. His research interests include statistical signal processing and in particular binaural noise reduction algorithms using microphone arrays.

Simon Doclo (S'95–M'03–SM'13) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in 1997 and 2003. From 2003 to 2007 he was a Postdoctoral Fellow with the Research Foundation Flanders at the Electrical Engineering Department (Katholieke Universiteit Leuven) and the Adaptive Systems Laboratory (McMaster University, Canada). From 2007 to 2009 he was a Principal Scientist with NXP Semiconductors at the Sound and Acoustics Group in Leuven, Belgium. Since 2009 he is a full professor at the University of Oldenburg, Germany, and scientific advisor for the project group Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research activities center around signal processing for acoustical applications, more specifically microphone array processing, active noise control, acoustic sensor networks and hearing aid processing. Prof. Doclo received the Master Thesis Award of the Royal Flemish Society of Engineers in 1997 (with Erik De Clippel), the Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2001, the EURASIP Signal Processing Best Paper Award in 2003 (with Marc Moonen) and the IEEE Signal Processing Society 2008 Best Paper Award (with Jingdong Chen, Jacob Benesty, Arden Huang).

He was member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2008–2013) and Technical Program Chair for the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2013. Prof. Doclo has served as guest editor for several special issues (IEEE Signal Processing Magazine, Elsevier Signal Processing) and is associate editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and EURASIP Journal on Advances in Signal Processing.

Sharon Gannot (S'92–M'01–SM'06) received his B.Sc. degree (summa cum laude) from the Technion Israel Institute of Technology, Haifa, Israel in 1986 and the M.Sc. (cum laude) and Ph.D. degrees from Tel-Aviv University, Israel in 1995 and 2000 respectively, all in Electrical Engineering. In 2001 he held a post-doctoral position at the department of Electrical Engineering (ESAT-SISTA) at K.U.Leuven, Belgium. From 2002 to 2003 he held a research and teaching position at the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. Currently, he is a Full Professor at the Faculty of Engineering, Bar-Ilan University, Israel, where he is heading the Speech and Signal Processing laboratory and the Signal Processing Track. Prof. Gannot is the recipient of Bar-Ilan University outstanding lecturer award for 2010 and 2014. Prof. Gannot has served as an Associate Editor of the EURASIP Journal of Advances in Signal Processing in 2003–2012, and as an Editor of several special issues on Multi-microphone Speech Processing of the same journal. He has also served as a guest editor of ELSEVIER Speech Communication and Signal Processing journals. Prof. Gannot has served as an Associate Editor of IEEE TRANSACTIONS ON SPEECH, AUDIO AND LANGUAGE PROCESSING in 2009–2013. Currently, he is a Senior Area Chair of the same journal. He also serves as a reviewer of many IEEE journals and conferences. Prof. Gannot is a member of the Audio and Acoustic Signal Processing (AASP) technical committee of the IEEE since Jan., 2010. Currently, he serves as the committee vice-chair. He is also a member of the Technical and Steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) since 2005 and was the general co-chair of IWAENC held at Tel-Aviv, Israel in August 2010. Prof. Gannot has served as the general co-chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in October 2013. Prof. Gannot was selected (with colleagues) to present a tutorial sessions in ICASSP 2012, EUSIPCO 2012, ICASSP 2013 and EUSIPCO 2013. Prof. Gannot research interests include multi-microphone speech processing and specifically distributed algorithms for ad hoc microphone arrays for noise reduction and speaker separation; dereverberation; single microphone speech enhancement and speaker localization and tracking.