

Binaural LCMV Beamforming With Partial Noise Estimation

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

Abstract—Besides reducing undesired sources, i.e., interfering sources and background noise, another important objective of a binaural beamforming algorithm is to preserve the spatial impression of the acoustic scene, which can be achieved by preserving the binaural cues of all sound sources. While the binaural minimum variance distortionless response (BMVDR) beamformer provides a good noise reduction performance and preserves the binaural cues of the desired source, it does not allow to control the reduction of the interfering sources and the background noise. Hence, several extensions have been proposed. First, the binaural linearly constrained minimum variance (BLCMV) beamformer uses additional constraints, enabling to control the reduction of the interfering sources while preserving their binaural cues. Second, the BMVDR with partial noise estimation (BMVDR-N) mixes the output signals of the BMVDR with the noisy reference microphone signals, enabling to control the binaural cues of the background noise. Aiming at merging the advantages of both extensions, in this paper we propose the BLCMV with partial noise estimation (BLCMV-N). We show that the output signals of the BLCMV-N can be interpreted as a mixture between the noisy reference microphone signals and the output signals of a BLCMV using an adjusted interference scaling parameter. We provide a theoretical comparison between the BMVDR, the BLCMV, the BMVDR-N and the proposed BLCMV-N in terms of noise and interference reduction performance and binaural cue preservation. Experimental results using recorded signals as well as the results of a perceptual listening test show that the BLCMV-N is able to preserve the binaural cues of an interfering source (like the BLCMV), while enabling to trade off between noise reduction performance and binaural cue preservation of the background noise (like the BMVDR-N).

Index Terms—Binaural cues, binaural noise reduction, MVDR beamformer, LCMV beamformer, hearing devices.

I. INTRODUCTION

BEAMFORMING algorithms for head-mounted assistive hearing devices (e.g., hearing aids, earbuds and hearables)

Manuscript received May 21, 2020; revised August 16, 2020; accepted September 21, 2020. Date of publication October 29, 2020; date of current version November 9, 2020. This work was supported in part by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project ID 352015383 (SFB 1330 B2) and Project ID 390895286 (EXC 2177/1) and in part by the Israeli Ministry of Science and Technology, #88962, 2019. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Andy W H Khong. (*Corresponding author: Nico Gößling.*)

Nico Gößling and Simon Doclo are with the Department of Medical Physics and Acoustics and the Cluster of Excellence Hearing4all, University of Oldenburg, 26111 Oldenburg, Germany (e-mail: nico.goessling@uol.de; simon.doclo@uol.de).

Elior Hadad and Sharon 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).

Digital Object Identifier 10.1109/TASLP.2020.3034526

are crucial to improve speech quality and speech intelligibility in noisy acoustic environments. Assuming a binaural configuration where both devices exchange their microphone signals, the information captured by all microphones on both sides of the head can be exploited [1]–[3]. Besides reducing interfering sources (e.g., competing speakers) and background noise (e.g., diffuse babble noise), another important objective of a binaural beamforming algorithm is the preservation of the listener’s spatial impression of the acoustic scene. This can be achieved by preserving the binaural cues of all sound sources, i.e., the interaural level difference (ILD) and the interaural time difference (ITD) for coherent sources (desired source and interfering sources) and the interaural coherence (IC) for incoherent sound fields (background noise) [4]. Binaural cues play a major role for spatial perception, i.e., to localize sound sources and to determine the spatial width or diffuseness of a sound field [5], and are very important for speech intelligibility due to so-called binaural unmasking [6], [7].

Unlike monaural beamforming algorithms, binaural beamforming algorithms need to generate two output signals (i.e., one for each ear), hence typically processing all available microphone signals from both devices by two different spatial filters [8]–[19]. A frequently used binaural beamforming algorithm is the binaural minimum variance distortionless response (BMVDR) beamformer, which aims at minimizing the power spectral density (PSD) of the noise component in the output signals while preserving the desired source component in the reference microphone signals on the left and the right device [2], [3], [11]. While the BMVDR provides a good noise reduction performance and preserves the binaural cues of the desired source, it does not allow to control the reduction of the interfering sources and distorts the binaural cues of the undesired sources (interfering sources and background noise). More specifically, after applying the BMVDR the binaural cues of the undesired sources are equal to the binaural cues of the desired source, such that all sources are perceived as coming from the same direction, which is obviously undesired. Hence, several extensions of the BMVDR have been proposed. On the one hand, the binaural linearly constrained minimum variance (BLCMV) beamformer uses additional interference reduction constraints, enabling to control the reduction of the interfering sources while preserving the binaural cues of the interfering sources in addition to the desired source by means of interference scaling parameters [12], [14], [17], [20]. However, due to the additional constraints there are less degrees of freedom available for noise reduction, such that the noise reduction performance for the BLCMV is

lower than for the BMVDR. Furthermore, it is not possible to explicitly trade off between noise reduction performance and binaural cue preservation of the background noise. On the other hand, the BMVDR with partial noise estimation (BMVDR-N) aims for the noise component in the output signals to be equal to a scaled version of the noise component in the reference microphone signals while preserving the desired source component in the reference microphone signals [3], [10], [11], [16]. It has been shown that the output signals of the BMVDR-N can be interpreted as a mixture between the output signals of the BMVDR and the noisy reference microphone signals, i.e., the BMVDR-N provides a trade-off between noise reduction performance and binaural cue preservation of the background noise. While for (incoherent) background noise the BMVDR-N showed promising results [16], [21], the effect of partial noise estimation on a (coherent) interfering source strongly depends on the relative position of the interfering source to the desired source and is harder to control [11].

Aiming at merging the advantages of the BLCMV and the BMVDR-N, i.e., preserving the binaural cues of the interfering sources and controlling the reduction of the interfering sources as well as the binaural cues of the background noise, in this paper we propose the BLCMV with partial noise estimation (BLCMV-N). First, we derive two decompositions for the BLCMV-N which reveal differences and similarities between the BLCMV-N and the BLCMV. We show that the output signals of the BLCMV-N can be interpreted as a mixture between the noisy reference microphone signals and the output signals of a BLCMV using an adjusted interference scaling parameter. We then analytically derive the performance of the BLCMV-N in terms of noise and interference reduction performance and binaural cue preservation. We show that the output signal-to-noise ratio (SNR) of the BLCMV-N is smaller than or equal to the output SNR of the BLCMV and derive the optimal interference scaling parameter maximizing the output SNR of the BLCMV-N. The derived analytical expressions are first validated using measured anechoic acoustic transfer functions (ATFs). In addition, more realistic experiments are performed using recorded signals for a binaural hearing device in a reverberant cafeteria with one interfering source and multi-talker babble noise. Both the objective performance measures as well as the results of a perceptual listening test with 13 normal-hearing participants show that the proposed BLCMV-N is able to preserve the binaural cues and hence the spatial impression of the interfering source (like the BLCMV), while trading off between noise reduction performance and binaural cue preservation of the background noise (like the BMVDR-N).

The remainder of this paper is organized as follows. In Section II we introduce the considered binaural hearing device configuration and the used objective performance measures. In Section III we briefly review several binaural beamforming algorithms, namely the BMVDR, the BLCMV and the BMVDR-N. In Section IV we present the BLCMV-N and derive two decompositions. In Section V we provide a detailed theoretical analysis of the proposed BLCMV-N in terms of noise and interference reduction performance and binaural cue preservation. In Section VI we first validate the analytical expressions

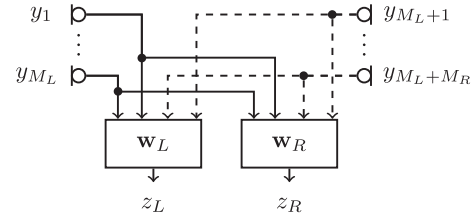


Fig. 1. Binaural hearing device configuration with M_L microphones on the left side and M_R microphones on the right side.

using anechoic ATFs, followed by simulations and a perceptual listening test using realistic recordings in a reverberant room.

II. HEARING DEVICE CONFIGURATION

In Section II-A the considered binaural hearing device configuration and the signal model are introduced. In Sections II-B and II-C the objective performance measures and the binaural cues are defined.

A. Signal Model

Consider the binaural hearing device configuration depicted in Fig. 1 with M_L microphones on the left side and M_R microphones on the right side, i.e., $M = M_L + M_R$ microphones in total. In this paper we consider an acoustic scenario with one desired source (target speaker) and one interfering source (competing speaker) in a noisy and reverberant environment, where the background noise is assumed to be incoherent (e.g., diffuse babble noise, sensor noise).

In the frequency-domain, the m -th microphone signal $y_m(\omega)$ can be decomposed as

$$y_m(\omega) = x_m(\omega) + u_m(\omega) + n_m(\omega) = x_m(\omega) + v_m(\omega), \quad (1)$$

with ω the normalized (radian) frequency, $x_m(\omega)$ the desired source component, $u_m(\omega)$ the interfering source component and $n_m(\omega)$ the noise component in the m -th microphone signal. The undesired component $v_m(\omega)$ is defined as the sum of the interfering source component $u_m(\omega)$ and the noise component $n_m(\omega)$. For the sake of conciseness, we omit the variable ω in the remainder of the paper wherever possible. The M -dimensional noisy input vector containing all microphone signals is defined as

$$\mathbf{y} = [y_1, \dots, y_{M_L}, y_{M_L+1}, \dots, y_M]^T, \quad (2)$$

where $(\cdot)^T$ denotes the transpose. Using (1), this vector can be written as

$$\mathbf{y} = \mathbf{x} + \mathbf{u} + \mathbf{n} = \mathbf{x} + \mathbf{v}, \quad (3)$$

where \mathbf{x} , \mathbf{u} , \mathbf{n} and \mathbf{v} are defined similarly as \mathbf{y} in (2).

For the considered acoustic scenario, the desired source component and the interfering source component can be written as

$$\mathbf{x} = s_x \mathbf{a}, \quad \mathbf{u} = s_u \mathbf{b}, \quad (4)$$

where s_x and s_u denote the desired source signal and the interfering source signal, respectively, and \mathbf{a} and \mathbf{b} denote M -dimensional ATF vectors, containing the ATFs between the

microphones and the desired source and the interfering source, respectively. It should be noted that the ATFs include reverberation, microphone characteristics and the head-shadow effect.

Without loss of generality, the first microphone on each side is defined as the so-called reference microphone. To simplify the notation, the reference microphone signals y_1 and y_{M_L+1} are denoted as y_L and y_R , i.e.,

$$y_L = \mathbf{e}_L^T \mathbf{y}, \quad y_R = \mathbf{e}_R^T \mathbf{y}, \quad (5)$$

where \mathbf{e}_L and \mathbf{e}_R denote M -dimensional selection vectors with all elements equal to 0 except one element equal to 1, i.e., $\mathbf{e}_L(1) = 1$ and $\mathbf{e}_R(M_L + 1) = 1$. Using (3), (4) and (5), the reference microphone signals can be written as

$$y_L = x_L + \underbrace{u_L + n_L}_{v_L} = a_L s_x + b_L s_u + n_L, \quad (6)$$

$$y_R = x_R + \underbrace{u_R + n_R}_{v_R} = a_R s_x + b_R s_u + n_R. \quad (7)$$

The noisy input covariance matrix \mathbf{R}_y , the desired source covariance matrix \mathbf{R}_x , the interfering source covariance matrix \mathbf{R}_u and the noise covariance matrix \mathbf{R}_n are defined as

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\}, \quad \mathbf{R}_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}, \quad (8)$$

$$\mathbf{R}_u = \mathcal{E}\{\mathbf{u}\mathbf{u}^H\}, \quad \mathbf{R}_n = \mathcal{E}\{\mathbf{n}\mathbf{n}^H\}, \quad (9)$$

with $\mathcal{E}\{\cdot\}$ the expected value operator and $(\cdot)^H$ the conjugate transpose. Assuming statistical independence between all signal components, \mathbf{R}_y can be written as

$$\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_u + \mathbf{R}_n = \mathbf{R}_x + \mathbf{R}_v, \quad (10)$$

with \mathbf{R}_v the undesired covariance matrix. Using (4), (8) and (9), the desired source covariance matrix and the interfering source covariance matrix can be written as rank-1 matrices, i.e.,

$$\mathbf{R}_x = p_x \mathbf{a}\mathbf{a}^H, \quad \mathbf{R}_u = p_u \mathbf{b}\mathbf{b}^H, \quad (11)$$

with $p_x = \mathcal{E}\{|s_x|^2\}$ the PSD of the desired source and $p_u = \mathcal{E}\{|s_u|^2\}$ the PSD of the interfering source. The noise covariance matrix \mathbf{R}_n is assumed to be full-rank, i.e., invertible and positive definite.

The left and the right output signals z_L and z_R are obtained by filtering and summing all microphone signals using the M -dimensional filter vectors \mathbf{w}_L and \mathbf{w}_R (cf. Fig. 1), i.e.,

$$z_L = \mathbf{w}_L^H \mathbf{y}, \quad z_R = \mathbf{w}_R^H \mathbf{y}. \quad (12)$$

B. Objective Performance Measures

The PSD and the cross power spectral density (CPSD) of the desired source component in the left and the right reference microphone signal are given by

$$p_{x,L}^{\text{in}} = \mathcal{E}\{|x_L|^2\} = \mathbf{e}_L^T \mathbf{R}_x \mathbf{e}_L = p_x |a_L|^2, \quad (13)$$

$$p_{x,R}^{\text{in}} = \mathcal{E}\{|x_R|^2\} = \mathbf{e}_R^T \mathbf{R}_x \mathbf{e}_R = p_x |a_R|^2, \quad (14)$$

$$p_{x,LR}^{\text{in}} = \mathcal{E}\{x_L x_R^*\} = \mathbf{e}_L^T \mathbf{R}_x \mathbf{e}_R = p_x a_L a_R^*. \quad (15)$$

Similarly, the output PSD of the desired source component in the left and the right output signal is given by

$$p_{x,L}^{\text{out}} = \mathbf{w}_L^H \mathbf{R}_x \mathbf{w}_L, \quad p_{x,R}^{\text{out}} = \mathbf{w}_R^H \mathbf{R}_x \mathbf{w}_R. \quad (16)$$

The same definitions can be applied for the noisy input signal, the interfering source component and the noise component by substituting \mathbf{R}_x with \mathbf{R}_y , \mathbf{R}_u or \mathbf{R}_n .

The narrowband input SNR in the left and the right reference microphone signal is defined as the ratio of the input PSD of the desired source and noise components, i.e.,

$$\text{SNR}_L^{\text{in}} = \frac{p_{x,L}^{\text{in}}}{p_{n,L}^{\text{in}}}, \quad \text{SNR}_R^{\text{in}} = \frac{p_{x,R}^{\text{in}}}{p_{n,R}^{\text{in}}}. \quad (17)$$

Similarly, the narrowband output SNR in the left and the right output signal is defined as the ratio of the output PSD of the desired source and noise components, i.e.,

$$\text{SNR}_L^{\text{out}} = \frac{p_{x,L}^{\text{out}}}{p_{n,L}^{\text{out}}}, \quad \text{SNR}_R^{\text{out}} = \frac{p_{x,R}^{\text{out}}}{p_{n,R}^{\text{out}}}. \quad (18)$$

The SNR improvement (in dB) is defined as $\Delta\text{SNR}_{L/R} = 10 \log_{10} \text{SNR}_{L/R}^{\text{out}} - 10 \log_{10} \text{SNR}_{L/R}^{\text{in}}$.

The narrowband input signal-to-interference ratio (SIR) in the left and the right reference microphone signal is defined as the ratio of the input PSD of the desired source and interfering source components, i.e.,

$$\text{SIR}_L^{\text{in}} = \frac{p_{x,L}^{\text{in}}}{p_{u,L}^{\text{in}}}, \quad \text{SIR}_R^{\text{in}} = \frac{p_{x,R}^{\text{in}}}{p_{u,R}^{\text{in}}}. \quad (19)$$

Similarly, the narrowband output SIR in the left and the right output signal is defined as the ratio of the output PSD of the desired source and interfering source components, i.e.,

$$\text{SIR}_L^{\text{out}} = \frac{p_{x,L}^{\text{out}}}{p_{u,L}^{\text{out}}}, \quad \text{SIR}_R^{\text{out}} = \frac{p_{x,R}^{\text{out}}}{p_{u,R}^{\text{out}}}. \quad (20)$$

The SIR improvement (in dB) is defined as $\Delta\text{SIR}_{L/R} = 10 \log_{10} \text{SIR}_{L/R}^{\text{out}} - 10 \log_{10} \text{SIR}_{L/R}^{\text{in}}$.

C. Binaural Cues

For coherent sources (desired source and interfering source) the main binaural cues used by the auditory system are the ILD and the ITD [4], which can be computed from the so-called interaural transfer function (ITF). Using (11), the input ITFs of the desired source and the interfering source are given by [11]

$$\text{ITF}_x^{\text{in}} = \frac{\mathcal{E}\{|x_L|^2\}}{\mathcal{E}\{x_R x_L^*\}} = \frac{\mathbf{e}_L^T \mathbf{R}_x \mathbf{e}_L}{\mathbf{e}_R^T \mathbf{R}_x \mathbf{e}_L} = \frac{a_L}{a_R}, \quad (21)$$

$$\text{ITF}_u^{\text{in}} = \frac{\mathcal{E}\{|u_L|^2\}}{\mathcal{E}\{u_R u_L^*\}} = \frac{\mathbf{e}_L^T \mathbf{R}_u \mathbf{e}_L}{\mathbf{e}_R^T \mathbf{R}_u \mathbf{e}_L} = \frac{b_L}{b_R}. \quad (22)$$

Similarly, the output ITFs of the desired source and the interfering source are given by

$$\text{ITF}_x^{\text{out}} = \frac{\mathbf{w}_L^H \mathbf{R}_x \mathbf{w}_L}{\mathbf{w}_R^H \mathbf{R}_x \mathbf{w}_L} = \frac{\mathbf{w}_L^H \mathbf{a}}{\mathbf{w}_R^H \mathbf{a}}, \quad (23)$$

$$\text{ITF}_u^{\text{out}} = \frac{\mathbf{w}_L^H \mathbf{R}_u \mathbf{w}_L}{\mathbf{w}_R^H \mathbf{R}_u \mathbf{w}_L} = \frac{\mathbf{w}_L^H \mathbf{b}}{\mathbf{w}_R^H \mathbf{b}}. \quad (24)$$

The ILD and the ITD can be calculated from the ITF as [11]

$$\text{ILD} = |\text{ITF}|^2, \quad \text{ITD} = \frac{\angle \text{ITF}}{\omega}, \quad (25)$$

with $\angle(\cdot)$ denoting the unwrapped phase.

For an incoherent sound field (background noise), ILD and ITD cues are not very descriptive, but the IC is known to play a major role for spatial perception (e.g., spatial width or diffuseness) [4]. The input IC of the noise component is defined as

$$\text{IC}_n^{\text{in}} = \frac{\mathcal{E}\{n_L n_R^*\}}{\sqrt{\mathcal{E}\{|n_L|^2\}}\sqrt{\mathcal{E}\{|n_R|^2\}}} = \frac{\mathbf{e}_L^T \mathbf{R}_n \mathbf{e}_R}{\sqrt{\mathbf{e}_L^T \mathbf{R}_n \mathbf{e}_L} \sqrt{\mathbf{e}_R^T \mathbf{R}_n \mathbf{e}_R}},$$

while the output IC of the noise component is defined as

$$\text{IC}_n^{\text{out}} = \frac{\mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_R}{\sqrt{\mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_L} \sqrt{\mathbf{w}_R^H \mathbf{R}_n \mathbf{w}_R}}. \quad (26)$$

Because the IC is typically complex-valued, the magnitude-squared coherence (MSC) is often used. The input and the output MSC of the noise component are defined as

$$\text{MSC}_n^{\text{in}} = |\text{IC}_n^{\text{in}}|^2, \quad \text{MSC}_n^{\text{out}} = |\text{IC}_n^{\text{out}}|^2. \quad (27)$$

An MSC of 1 corresponds to a coherent source perceived as a distinct point source, while smaller MSC values correspond to a broader or even diffuse sound field impression [4].

III. BINAURAL BEAMFORMING ALGORITHMS

In this section we briefly review three state-of-the-art binaural beamforming algorithms, namely the BMVDR beamformer, the BLCMV beamformer and the BMVDR-N beamformer. We discuss the performance of these beamforming algorithms in terms of noise and interference reduction performance and binaural cue preservation. For the sake of conciseness, we only show expressions for the left hearing device, denoted by the subscript L . It should be noted that all expressions can also be formulated for the right hearing device by changing the subscript to R .

A. BMVDR Beamformer

The BMVDR aims at minimizing the output PSD of the noise component while preserving the desired source component in the reference microphone signals [2], [3], [11]. The constrained optimization problem for the left filter vector is given by

$$\min_{\mathbf{w}_L} \mathcal{E}\{|\mathbf{w}_L^H \mathbf{n}|^2\} \quad \text{s.t.} \quad \mathbf{w}_L^H \mathbf{x} = x_L \quad (28)$$

Using (4), (6) and (9), the solution of (28) is equal to [2], [22], [23]

$$\mathbf{w}_{\text{BMVDR},L} = \frac{\mathbf{R}_n^{-1} \mathbf{a}}{\gamma_a} a_L^* \quad (29)$$

with

$$\gamma_a = \mathbf{a}^H \mathbf{R}_n^{-1} \mathbf{a}. \quad (30)$$

It should be noted that the BMVDR can also be defined using the undesired covariance matrix \mathbf{R}_v instead of the noise covariance matrix \mathbf{R}_n . However, since \mathbf{R}_v is considerably more difficult to estimate or model in practice than \mathbf{R}_n , in this paper we only consider the BMVDR using \mathbf{R}_n in (29).

By substituting (29) in (18) and (20), it has been shown in [3], [11] that the output SNR and the output SIR of the BMVDR are equal to

$$\text{SNR}_{\text{BMVDR},L}^{\text{out}} = p_x \gamma_a, \quad (31)$$

$$\text{SIR}_{\text{BMVDR},L}^{\text{out}} = \frac{p_x |\gamma_a|^2}{p_u |\gamma_{ab}|^2}, \quad (32)$$

with γ_a defined in (30) and

$$\gamma_{ab} = \mathbf{a}^H \mathbf{R}_n^{-1} \mathbf{b}. \quad (33)$$

Although the BMVDR yields the largest output SNR among all distortionless binaural beamforming algorithms, the output SIR depends on the relative position of the interfering source to the desired source, cf. (33).

As shown in [3], [11], [13], the BMVDR preserves the binaural cues of the desired source, but distorts the binaural cues of the undesired sources. Hence, at the output of the BMVDR the interfering source and the (incoherent) background noise are perceived as coming from the direction of the desired source, which is obviously undesired in terms of spatial awareness.

B. BLCMV Beamformer

In addition to preserving the desired source component in the reference microphone signals, the BLCMV preserves a scaled version of the interfering source component in the reference microphone signals while minimizing the output PSD of the noise component [12], [14]. The constrained optimization problem for the left filter vector is given by [14]

$$\min_{\mathbf{w}_L} \mathcal{E}\{|\mathbf{w}_L^H \mathbf{n}|^2\} \quad \text{s.t.} \quad \mathbf{w}_L^H \mathbf{x} = x_L, \quad \mathbf{w}_L^H \mathbf{u} = \delta u_L \quad (34)$$

with $0 < \delta \leq 1$ the (real-valued) interference scaling parameter. Using (4), (6) and (9), the solution of (34) is equal to [14]

$$\mathbf{w}_{\text{BLCMV},L} = \mathbf{R}_n^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}_n^{-1} \mathbf{C})^{-1} \mathbf{g}_L \quad (35)$$

with the constraint matrix \mathbf{C} and the left response vector \mathbf{g}_L defined as

$$\mathbf{C} = [\mathbf{a}, \mathbf{b}], \quad \mathbf{g}_L = [a_L, \delta b_L]^H. \quad (36)$$

By substituting (35) in (18), it has been shown in [14] that the output SNR of the BLCMV is equal to

$$\text{SNR}_{\text{BLCMV},L}^{\text{out}} = \frac{p_x |a_L|^2}{\mathbf{e}_L^T \mathbf{R}_{x_u,1} \mathbf{e}_L}, \quad (37)$$

with

$$\mathbf{R}_{x_u,1} = \frac{1}{1 - \Psi} \left[\frac{\mathbf{a}\mathbf{a}^H}{\gamma_a} + \delta^2 \frac{\mathbf{b}\mathbf{b}^H}{\gamma_b} - 2\Psi\delta\Re \left\{ \frac{\mathbf{a}\mathbf{b}^H}{\gamma_{ab}} \right\} \right], \quad (38)$$

$$\gamma_b = \mathbf{b}^H \mathbf{R}_n^{-1} \mathbf{b}, \quad \Psi = \frac{|\gamma_{ab}|^2}{\gamma_a \gamma_b}, \quad (39)$$

where $\Re\{\cdot\}$ denotes the real part of a complex number. The output SNR of the BLCMV in (37) is smaller than or equal to the output SNR of the BMVDR in (31), since less degrees of

freedom are available for noise reduction. In addition, the output SIR of the BLCMV is equal to [14]

$$\text{SIR}_{\text{BLCMV},L}^{\text{out}} = \frac{1}{\delta^2} \text{SIR}_L^{\text{in}}, \quad (40)$$

which can hence be directly controlled by the interference scaling parameter δ .

As shown in [14], the BLCMV preserves the binaural cues of both the desired source and the interfering source and the output MSC of the noise component is equal to

$$\text{MSC}_{\text{BLCMV},n}^{\text{out}} = \left| \frac{\mathbf{e}_L^T \mathbf{R}_{xu,1} \mathbf{e}_R}{\sqrt{\mathbf{e}_L^T \mathbf{R}_{xu,1} \mathbf{e}_L} \sqrt{\mathbf{e}_R^T \mathbf{R}_{xu,1} \mathbf{e}_R}} \right|^2. \quad (41)$$

Because $\mathbf{R}_{xu,1}$ in (38) is a rank-2 matrix, it has been shown in [14] that the output MSC of the noise component is smaller than 1 but is not equal to the input MSC of the noise component. Furthermore, it should be noted that the output MSC of the noise component depends on the relative position of the interfering source to the desired source, cf. (38) and (39), such that it is not straightforward to control the binaural cues of the background noise.

C. BMVDR-N Beamformer

In addition to preserving the desired source component in the reference microphone signals, the BMVDR with partial noise estimation (BMVDR-N) aims at preserving a scaled version of the noise component in the reference microphone signals [3], [10], [11]. The constrained optimization problem for the left filter vector is given by

$$\min_{\mathbf{w}_L} \mathcal{E} \left\{ |\mathbf{w}_L^H \mathbf{n} - \eta n_L|^2 \right\} \quad \text{s.t.} \quad \mathbf{w}_L^H \mathbf{x} = x_L \quad (42)$$

with $0 \leq \eta \leq 1$ the (real-valued) mixing parameter. It has been shown in [11] that the solution of (42) is equal to

$$\mathbf{w}_{\text{BMVDR-N},L} = (1 - \eta) \mathbf{w}_{\text{BMVDR},L} + \eta \mathbf{e}_L \quad (43)$$

with $\mathbf{w}_{\text{BMVDR},L}$ defined in (29). Hence, the output signals of the BMVDR-N can be interpreted as a mixture between the noisy reference microphone signals (scaled with η) and the output signals of the BMVDR (scaled with $1 - \eta$). For $\eta = 0$, the BMVDR-N is equal to the BMVDR, whereas for $\eta = 1$, no beamforming is applied.

Since the output signals of the BMVDR are mixed with the noisy reference microphone signals, the output SNR of the BMVDR-N is always smaller than or equal to the output SNR of the BMVDR [11], i.e.,

$$\text{SNR}_{\text{BMVDR-N},L}^{\text{out}} \leq \text{SNR}_{\text{BMVDR},L}^{\text{out}} \quad (44)$$

and decreases with increasing η . By substituting (43) in (20), it can be shown that the output SIR of the BMVDR-N is equal to

$$\text{SIR}_{\text{BMVDR-N},L}^{\text{out}} = \frac{p_x}{p_u} \frac{|a_L|^2}{\mathbf{e}_L^T \mathbf{R}_{xu,2} \mathbf{e}_L}, \quad (45)$$

with

$$\begin{aligned} \mathbf{R}_{xu,2} &= (1 - \eta)^2 \frac{|\gamma_{ab}|^2}{|\gamma_a|^2} \mathbf{a} \mathbf{a}^H \\ &+ \eta^2 \mathbf{b} \mathbf{b}^H + (\eta - \eta^2) 2\Re \left\{ \mathbf{a} \mathbf{b}^H \frac{\gamma_{ab}}{\gamma_a} \right\}. \end{aligned}$$

As shown in [11], [16], the BMVDR-N preserves the binaural cues of the desired source. It has further been shown in [16] and [20] that only for $\eta = 1$ the binaural cues of the undesired sources (interfering source and background noise) are preserved, whereas for $\eta = 0$ the binaural cues of the undesired sources are equal to the binaural cues of the desired source (as for the BMVDR). The mixing parameter η hence allows to trade off between noise reduction performance and binaural cue preservation of the background noise, or in other words control the binaural cues of the background noise. Furthermore, it should be noted that the interference reduction performance in (45) and the output ITF of the interfering source do not only depend on the mixing parameter η but also on the relative position of the interfering source to the desired source, such that it is not straightforward to control both.

IV. BLCMV WITH PARTIAL NOISE ESTIMATION

Aiming at merging the advantages of the BLCMV and the BMVDR-N, i.e., preserving the binaural cues of the interfering source and controlling the binaural cues of the background noise, in Section IV-A we present the BLCMV beamformer with partial noise estimation (BLCMV-N). Similarly as for the BLCMV in [14], in Sections IV-B and IV-C we derive two decompositions for the BLCMV-N which reveal differences and similarities between the BLCMV-N and the BLCMV.

A. BLCMV-N Beamformer

Compared to the BMVDR in (28), the BLCMV-N uses an additional constraint to preserve a scaled version of the interfering source component in the reference microphone signals, like the BLCMV in (34), and aims at preserving a scaled version of the noise component in the reference microphone signals, like the BMVDR-N in (42). The constrained optimization problem for the left filter vector is given by

$$\min_{\mathbf{w}_L} \mathcal{E} \left\{ |\mathbf{w}_L^H \mathbf{n} - \eta n_L|^2 \right\} \quad \text{s.t.} \quad \mathbf{w}_L^H \mathbf{x} = x_L, \quad \mathbf{w}_L^H \mathbf{u} = \delta u_L \quad (46)$$

The solution of (46) is equal to (see Appendix A)

$$\mathbf{w}_{\text{BLCMV-N},L} = \eta \mathbf{e}_L + (1 - \eta) \mathbf{R}_n^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}_n^{-1} \mathbf{C})^{-1} \begin{bmatrix} a_L^* \\ \bar{\delta} b_L^* \end{bmatrix} \quad (47)$$

with \mathbf{C} defined in (36) and the *adjusted interference scaling parameter* $\bar{\delta}$ equal to

$$\bar{\delta} = \frac{\delta - \eta}{1 - \eta}. \quad (48)$$

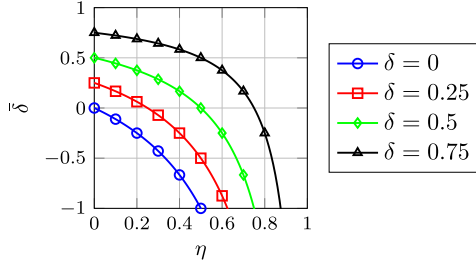


Fig. 2. Adjusted interference scaling parameter $\bar{\delta}$ as a function of η for different values of δ .

Hence, the output signals of the BLCMV-N can be interpreted as a mixture between the noisy reference microphone signals (scaled with η) and the output signals of a BLCMV (scaled with $1 - \eta$) using the adjusted interference scaling parameter $\bar{\delta}$ in (48) instead of the interference scaling parameter δ . For $\eta = 0$, the BLCMV-N is equal to the BLCMV in (35) with $\bar{\delta} = \delta$, whereas for $\eta = 1$, it should be realized that only if $\delta = 1$ no beamforming is applied. Since mixing with the reference microphone signals not only affects the noise component but also the interfering source component, the adjusted interference scaling parameter $\bar{\delta}$ depends on both the interference scaling parameter δ as well as the mixing parameter η due to the interference reduction constraint in (46). Fig. 2 depicts $\bar{\delta}$ as a function of η for different values of δ . It can be seen that

$$\bar{\delta}(\eta, \delta) = \begin{cases} > 0, & \text{for } \delta > \eta \\ < 0, & \text{for } \delta < \eta \\ 0, & \text{for } \delta = \eta \end{cases} \quad (49)$$

As will be shown in more detail in the following sections, using the parameters δ and η it is possible to control the noise reduction performance, the interference reduction performance and the binaural cues of the background noise for the BLCMV-N.

B. Decomposition Into Two BLCMVs

In [14] it has been shown that the BLCMV in (35) can be decomposed as the sum of two sub-BLCMVs, i.e.,

$$\mathbf{w}_{\text{BLCMV},L} = \mathbf{w}_{x,L} + \delta \mathbf{w}_{u,L}, \quad (50)$$

with

$$\mathbf{w}_{x,L} = \mathbf{R}_n^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}_n^{-1} \mathbf{C})^{-1} \mathbf{g}_{x,L}, \quad (51)$$

$$\mathbf{w}_{u,L} = \mathbf{R}_n^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}_n^{-1} \mathbf{C})^{-1} \mathbf{g}_{u,L}, \quad (52)$$

and the respective response vectors

$$\mathbf{g}_{x,L} = \begin{bmatrix} a_L^* \\ 0 \end{bmatrix}, \quad \mathbf{g}_{u,L} = \begin{bmatrix} 0 \\ b_L^* \end{bmatrix}. \quad (53)$$

The sub-BLCMV $\mathbf{w}_{x,L}$ in (51) preserves the desired source component in the reference microphone signals and steers a null towards the interfering source, whereas the sub-BLCMV $\mathbf{w}_{u,L}$ in (52) preserves the interfering source component in the reference microphone signals and steers a null towards the desired source. Using (47), it can be easily seen that the proposed

BLCMV-N can be decomposed as

$$\mathbf{w}_{\text{BLCMV-N},L} = \eta \mathbf{e}_L + (1 - \eta) \mathbf{w}_{x,L} + (\delta - \eta) \mathbf{w}_{u,L} \quad (54)$$

Hence, the BLCMV-N can be interpreted as a mixture of the reference microphone signals (scaled with η), a BLCMV that preserves the desired source and rejects the interfering source (scaled with $1 - \eta$) and a BLCMV that preserves the interfering source and rejects the desired source (scaled with $\delta - \eta$). Since the scaling of the sub-BLCMV $\mathbf{w}_{x,L}$ controls the desired source component without affecting the interfering source component and the scaling of the sub-BLCMV $\mathbf{w}_{u,L}$ controls the interfering source component without affecting the desired source component [14], it can be directly observed from the scaling factors in (54) that the desired source component is not distorted and the interfering source component is scaled with δ .

C. Decomposition Using Binauralization Postfilters

In [14] it has also been shown that the sub-BLCMV $\mathbf{w}_{x,L}$ in (51) for the left hearing device and the sub-BLCMV $\mathbf{w}_{x,R}$ for the right hearing device (defined similarly as $\mathbf{w}_{x,L}$) can be written using a common spatial filter and two binauralization postfilters as

$$\mathbf{w}_{x,L} = \mathbf{w}_x a_L^*, \quad \mathbf{w}_{x,R} = \mathbf{w}_x a_R^*, \quad (55)$$

with the common desired BLCMV (D-BLCMV) given by

$$\mathbf{w}_x = \frac{1}{1 - \Psi} \left(\frac{\mathbf{R}_n^{-1} \mathbf{a}}{\gamma_a} - \Psi \frac{\mathbf{R}_n^{-1} \mathbf{b}}{\gamma_{ab}} \right), \quad (56)$$

and the ATFs a_L and a_R between the desired source and the reference microphones used as binauralization postfilters. Similarly, the sub-BLCMV $\mathbf{w}_{u,L}$ in (52) and the sub-BLCMV $\mathbf{w}_{u,R}$ (defined similarly as $\mathbf{w}_{u,L}$) can be written as

$$\mathbf{w}_{u,L} = \mathbf{w}_u b_L^*, \quad \mathbf{w}_{u,R} = \mathbf{w}_u b_R^*, \quad (57)$$

with the common interference BLCMV (I-BLCMV) given by

$$\mathbf{w}_u = \frac{1}{1 - \Psi} \left(\frac{\mathbf{R}_n^{-1} \mathbf{b}}{\gamma_b} - \Psi \frac{\mathbf{R}_n^{-1} \mathbf{a}}{\gamma_{ab}^*} \right), \quad (58)$$

and the ATFs b_L and b_R between the interfering source and the reference microphones used as binauralization postfilters.

Using (55) and (57) in (54), the BLCMV-N can be decomposed as

$$\mathbf{w}_{\text{BLCMV-N},L} = \eta \mathbf{e}_L + (1 - \eta) a_L^* \mathbf{w}_x + (\delta - \eta) b_L^* \mathbf{w}_u \quad (59)$$

$$\mathbf{w}_{\text{BLCMV-N},R} = \eta \mathbf{e}_R + (1 - \eta) a_R^* \mathbf{w}_x + (\delta - \eta) b_R^* \mathbf{w}_u \quad (60)$$

Fig. 3 depicts this decomposition of the BLCMV-N using common spatial filters and binauralization postfilters. The output signals of the BLCMV-N can hence be interpreted as a mixture between the reference microphone signals (scaled with η), the binauralized output signals of the D-BLCMV (scaled with $1 - \eta$) and the binauralized output signals of the I-BLCMV (scaled with $\delta - \eta$).

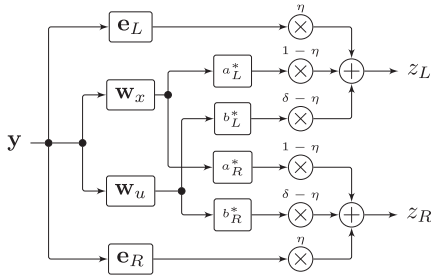


Fig. 3. Decomposition of the BLCMV-N into a mixture of the reference microphone signals and two BLCMVs with binauralization postfilters.

Due to the constraints in (46), the BLCMV-N perfectly preserves the desired source component and scales the interfering source component with δ . Using (59) and (60), the noise component in the output signals of the BLCMV-N are equal to

$$\mathbf{w}_{\text{BLCMV-N},L}^H \mathbf{n} = \eta n_L + (1 - \eta) n_x a_L + (\delta - \eta) n_u b_L, \quad (61)$$

$$\mathbf{w}_{\text{BLCMV-N},R}^H \mathbf{n} = \eta n_R + (1 - \eta) n_x a_R + (\delta - \eta) n_u b_R, \quad (62)$$

with $n_x = \mathbf{w}_x^H \mathbf{n}$ and $n_u = \mathbf{w}_u^H \mathbf{n}$ the noise component in the output signal of the D-BLCMV and the I-BLCMV, respectively. The noise component in the output signals of the BLCMV-N can hence be interpreted as a mixture between the noise component in the reference microphone signals (scaled with η), a coherent residual noise source (n_x) coming from the direction of the desired source (scaled with $1 - \eta$) and a coherent residual noise source (n_u) coming from the direction of the interfering source (scaled with $\delta - \eta$).

V. PERFORMANCE OF THE BLCMV-N

In this section we provide a performance analysis of the proposed BLCMV-N. In Section V-A we derive the output PSDs of the signal components. In Sections V-B and V-C we analyze the noise and interference reduction performance and the binaural cue preservation performance. Finally, in Section V-D we discuss the setting of the mixing parameter η and the interference scaling parameter δ .

A. Output Power Spectral Densities

Due to the constraints in (46), the output PSD of the desired and interfering source components in the left output signal of the BLCMV-N are equal to, cf. (13),

$$p_{\text{BLCMV-N},x,L}^{\text{out}} = p_{x,L}^{\text{in}} = p_x |a_L|^2, \quad (63)$$

$$p_{\text{BLCMV-N},u,L}^{\text{out}} = \delta^2 p_{u,L}^{\text{in}} = \delta^2 p_u |b_L|^2. \quad (64)$$

Furthermore, the output PSD of the noise component in the left output signal of the BLCMV-N is equal to (see Appendix B)

$$p_{\text{BLCMV-N},n,L}^{\text{out}} = \mathbf{e}_L^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{xu,3}) \mathbf{e}_L, \quad (65)$$

with

$$\mathbf{R}_{xu,3} = \frac{1}{1 - \Psi} \left[(1 - \eta^2) \frac{\mathbf{a}\mathbf{a}^H}{\gamma_a} + (\delta^2 - \eta^2) \frac{\mathbf{b}\mathbf{b}^H}{\gamma_b} - 2\Psi(\delta - \eta^2) \Re \left\{ \frac{\mathbf{a}\mathbf{b}^H}{\gamma_{ab}^*} \right\} \right], \quad (66)$$

with γ_a defined in (30), γ_{ab} defined in (33), and γ_b and Ψ defined in (39). It can be seen that the output PSD of the noise component for the BLCMV-N is a quadratic function in both the mixing parameter η and the interference scaling parameter δ . By comparing (66) to (38), it can be observed that

$$\mathbf{R}_{xu,3} = \mathbf{R}_{xu,1} - \eta^2 \mathbf{R}_{xu,1}^{\delta=1} \quad (67)$$

where $\mathbf{R}_{xu,1}^{\delta=1}$ denotes the expression for the BLCMV in (38) with $\delta = 1$, corresponding to no suppression of the interfering source. Please note that for $\eta = 0$, $\mathbf{R}_{xu,3} = \mathbf{R}_{xu,1}$, and for $\eta = 1$ and $\delta = 1$, $\mathbf{R}_{xu,3} = \mathbf{0}_M$. By using (67) in (65), it follows that

$$p_{\text{BLCMV-N},n,L}^{\text{out}} = \eta^2 \left(p_{n,L}^{\text{in}} - p_{\text{BLCMV},n,L}^{\text{out},\delta=1} \right) + p_{\text{BLCMV},n,L}^{\text{out}}. \quad (68)$$

B. Noise and Interference Reduction Performance

By substituting (63) and (65) in (18), the left output SNR of the BLCMV-N is equal to

$$\text{SNR}_{\text{BLCMV-N},L}^{\text{out}} = \frac{p_x |a_L|^2}{\mathbf{e}_L^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{xu,3}) \mathbf{e}_L}, \quad (69)$$

which depends on both the mixing parameter η and the interference scaling parameter δ . Using (68) and realizing that the output PSD of the noise component in the left output signal of the BLCMV (for any value of δ) is smaller than or equal to the PSD of the noise component in the left reference microphone signal, the output SNR of the BLCMV-N in (69) is smaller than or equal to the output SNR of the BLCMV in (37), i.e.,

$$\text{SNR}_{\text{BLCMV-N},L}^{\text{out}} \leq \text{SNR}_{\text{BLCMV},L}^{\text{out}} \leq \text{SNR}_{\text{BMVDR},L}^{\text{out}} \quad (70)$$

By substituting (63) and (64) in (20), the left output SIR of the BLCMV-N is equal to

$$\text{SIR}_{\text{BLCMV-N},L}^{\text{out}} = \frac{1}{\delta^2} \text{SIR}_L^{\text{in}}, \quad (71)$$

which is equal to the left output SIR of the BLCMV in (40) and solely controlled by the interference scaling parameter δ . For $\eta = 0$, the left output SNR of the BLCMV-N is equal to the left output SNR of the BLCMV in (37), while for $\eta = 1$ and $\delta = 1$, the left output SNR of the BLCMV-N is equal to the left input SNR because no beamforming is applied.

C. Binaural Cue Preservation

Similarly as for the BLCMV, due to the constraints in (46) the BLCMV-N preserves the binaural cues of both the desired source and the interfering source, i.e.,

$$\text{ITF}_{\text{BLCMV-N},x}^{\text{out}} = \frac{a_L}{a_R} = \text{ITF}_x^{\text{in}}, \quad (72)$$

$$\text{ITF}_{\text{BLCMV-N},u}^{\text{out}} = \frac{b_L}{b_R} = \text{ITF}_u^{\text{in}}. \quad (73)$$

TABLE I
NOISE AND INTERFERENCE REDUCTION PERFORMANCE AND BINAURAL CUE PRESERVATION OF ALL CONSIDERED BINAURAL BEAMFORMING ALGORITHMS.
†: DEPENDS ON RELATIVE POSITION OF INTERFERING SOURCE TO DESIRED SOURCE

Algorithm	SNR	SIR	ITF _x	ITF _u	MSC _n
BMVDR	+++	†	preserved	not preserved	not preserved
BLCMV	++	controlled by δ	preserved	preserved	not preserved, †
BMVDR-N	++	†	preserved	not preserved, †	controlled by η
BLCMV-N	+	controlled by δ	preserved	preserved	controlled by η

Using (26), the output IC of the noise component for the BLCMV-N is equal to (see Appendix B for derivation of components)

$$\begin{aligned} & \text{IC}_{\text{BLCMV-N},n}^{\text{out}} \\ &= \frac{\mathbf{e}_L^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{x_{u,3}}) \mathbf{e}_R}{\sqrt{\mathbf{e}_L^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{x_{u,3}}) \mathbf{e}_L} \sqrt{\mathbf{e}_R^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{x_{u,3}}) \mathbf{e}_R}}, \end{aligned} \quad (74)$$

with $\mathbf{R}_{x_{u,3}}$ defined in (66). Since $\mathbf{R}_{x_{u,3}}$ depends on both the mixing parameter η and the interference scaling parameter δ , also the output IC of the noise component in (74) depends on both parameters. Using (27), the output MSC of the noise component for the BLCMV-N is equal to

$$\text{MSC}_{\text{BLCMV-N},n}^{\text{out}} = |\text{IC}_{\text{BLCMV-N},n}^{\text{out}}|^2. \quad (75)$$

Since for $\eta = 0$ the BLCMV-N is equal to the BLCMV, the output MSC of the noise component is smaller than 1, see Section III-B. It should however be realized that in contrast to the BMVDR-N discussed in Section III-C, for $\eta = 1$ the BLCMV-N does not always preserve the MSC of the noise component. Only for $\eta = 1$ and $\delta = 1$ the binaural cues of all signal components are preserved because no beamforming is applied.

Table I summarizes the noise and interference reduction performance and binaural cue preservation of all considered binaural beamforming algorithms.

D. Parameter Settings

Maximizing the left output SNR in (69) corresponds to minimizing the denominator, i.e., using (67),

$$D(\eta, \delta) = \mathbf{e}_L^T [\eta^2 (\mathbf{R}_n - \mathbf{R}_{x_{u,1}}^{\delta=1}) + \mathbf{R}_{x_{u,1}}] \mathbf{e}_L. \quad (76)$$

Setting the derivative of (76) with respect to the mixing parameter η equal to zero, yields

$$\eta_{\text{opt}} = 0 \quad (77)$$

as the optimal mixing parameter η in terms of left (and right) output SNR. However, it should be noted again that for $\eta = 0$ the BLCMV-N is equal to the BLCMV and the output MSC of the noise component depends on the relative position of the interfering source to the desired source, such that it is not straightforward to control the binaural cues of the background noise. Similarly as for the BMVDR-N, cf. [11], [16] and simulations in Section VI, setting the mixing parameter η to a value between 0.2 and 0.3 seems to be a good compromise between noise reduction

performance and binaural cue preservation of the background noise in practice. As shown in [13], [16] for the BMVDR-N, the mixing parameter η may also be psycho-acoustically motivated based on the binaural cue discrimination ability of the human auditory system.

The derivative of (76) with respect to the interference scaling parameter δ is equal to, using (38),

$$\frac{\partial D(\eta, \delta)}{\partial \delta} = \frac{1}{1 - \Psi} \left(2\delta \frac{|b_L|^2}{\gamma_b} - 2\Psi \Re \left\{ \frac{a_L b_L^*}{\gamma_{ab}^*} \right\} \right). \quad (78)$$

Setting (78) to zero and solving for δ yields the optimal interference scaling parameter in terms of left output SNR, i.e.,

$$\delta_{\text{opt},L} = \frac{\alpha_L}{\beta_L}, \quad (79)$$

with

$$\alpha_L = \Psi \Re \left\{ \frac{a_L b_L^*}{\gamma_{ab}^*} \right\}, \quad \beta_L = \frac{|b_L|^2}{\gamma_b}, \quad (80)$$

which depends on the relative position of the interfering source to the desired source. As can be seen from (71), the output SIR is not affected by the mixing parameter η but is solely determined by the interference scaling parameter δ . Similarly as experimentally shown for the BLCMV in [17], setting the interference scaling parameter to $\delta = 0$ maximizes the output SIR of the BLCMV-N (cf. (71)), but may lead to robustness problems in practice due to estimation errors. Hence, it is recommended to set the interference scaling parameter δ to a value between 0.2 and 0.3, which seems to be good compromise between interference reduction and robustness in practice (cf. simulations in Section VI).

VI. SIMULATIONS

In Section VI-A we first validate the expressions derived in the previous sections using measured anechoic ATFs. In Section VI-B we then experimentally compare the performance of the proposed BLCMV-N with the BMVDR, BLCMV and BMVDR-N using recorded signals in a reverberant environment with a competing speaker and multi-talker babble noise. Finally, in Section VI-C we compare the spatial impression of the considered binaural beamforming algorithms using a perceptual listening test.

A. Validation Using Measured Anechoic ATFs

To validate the derived expressions for the considered algorithms we used measured anechoic ATFs of two behind-the-ear hearing aids mounted on a head-and-torso-simulator (HATS) [24]. Each hearing aid has two microphones ($M = 4$) with an inter-microphone distance of about 14 mm. We chose the front microphone on each hearing aid as reference microphone. The ATFs were calculated from anechoic impulse responses using a 512-point FFT at a sampling rate of 16 kHz.

The desired source was placed at 0° (in front) and the interfering source was placed at -35° (to the left), both at a distance of 3 m from the HATS. The desired source covariance matrix \mathbf{R}_x and the interfering source covariance matrix \mathbf{R}_u were constructed using the ATF vector of the desired source \mathbf{a} and the ATF vector of the interfering source \mathbf{b} according to (11), where the PSD

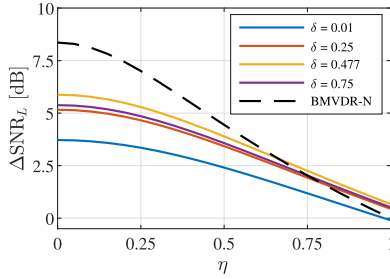


Fig. 4. SNR improvement for the BLCMV-N and the BMVDR-N at 500 Hz.

of the desired source p_x and the PSD of the interfering source p_u were both set to 1. As background noise we considered a combination of spatially white and cylindrically isotropic noise, i.e., the noise covariance matrix \mathbf{R}_n was constructed as

$$\mathbf{R}_n = p_{n,w}\mathbf{I}_M + p_{n,cyl}\mathbf{\Gamma}, \quad (81)$$

with $p_{n,w}$ the PSD of the spatially white noise, \mathbf{I}_M the $M \times M$ -dimensional identity matrix, $p_{n,cyl}$ the PSD of the cylindrically isotropic noise and $\mathbf{\Gamma}$ its spatial coherence matrix. The (i, j) -th element of the spatial coherence matrix $\mathbf{\Gamma}$ was calculated using all available anechoic ATF's as

$$\Gamma_{i,j} = \frac{\sum_{k=1}^K h_i(\theta_k)h_j^*(\theta_k)}{\sqrt{\sum_{k=1}^K |h_i(\theta_k)|^2}\sqrt{\sum_{k=1}^K |h_j(\theta_k)|^2}}, \quad (82)$$

with $h(\theta_k)$ the anechoic ATF at angle θ_k and K the total number of angles in the database ($K = 72$ for [24]). The PSD of the spatially white noise $p_{n,w}$ was set to -55 dB, while the PSD of the cylindrically isotropic noise $p_{n,cyl}$ was set to 1.

1) *Noise and Interference Reduction Performance:* Using (17) and (18), Fig. 4 depicts the left SNR improvement at 500 Hz for the BLCMV-N for different values of the mixing parameter η and the interference scaling parameter δ and the BMVDR-N for different values of the mixing parameter η . As expected, the BMVDR (i.e., BMVDR-N for $\eta = 0$) yields the largest SNR improvement (cf. (70)). Since the BMVDR-N mixes the output signals of the BMVDR with the noisy reference microphone signals, it can be observed that increasing the mixing parameter η reduces the SNR improvement of the BMVDR-N compared to the BMVDR ($\eta = 0$). For the BLCMV-N, both η and δ affect the SNR improvement, which is in line with (69). Similarly to the BMVDR-N, the BLCMV-N mixes the output signals of a BLCMV with the noisy reference microphone signals. Hence, it can be observed that for any value of the interference scaling parameter δ , increasing the mixing parameter η reduces the SNR improvement of the BLCMV-N compared to the BLCMV ($\eta = 0$), which is in line with (70). Since less degrees of freedom are available for noise reduction, the BLCMV ($\eta = 0$) yields a smaller SNR improvement compared to the BMVDR ($\eta = 0$), as discussed in Section III-B. Using (79), the interference scaling parameter δ maximizing the output SNR was equal to $\delta_{\text{opt},L} = 0.477$ for the considered acoustic scenario. As expected, it can be observed that using $\delta_{\text{opt},L}$ leads to the largest SNR improvement of all considered values of δ . For large values of the mixing parameter η , the BLCMV-N yields a larger SNR improvement than the

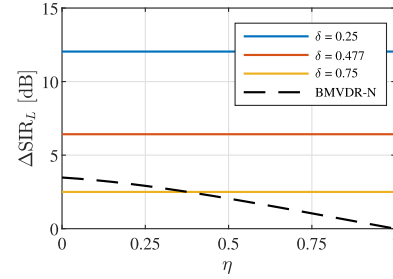


Fig. 5. SIR improvement for the BLCMV-N and the BMVDR-N at 500 Hz.

BMVDR-N. It should be noted that the exact behaviour depends on the interference scaling parameter δ and the relative position of the interfering source to the desired source.

Using (19) and (20), Fig. 5 depicts the left SIR improvement at 500 Hz for the BLCMV-N for different values of the mixing parameter η and the interference scaling parameter δ and the BMVDR-N for different values of the mixing parameter η . As expected from (40) and (71), both the BLCMV-N and the BLCMV ($\eta = 0$) yield the same SIR improvement, which is solely controlled by the interference scaling parameter δ . Hence, increasing the interference scaling parameter δ reduces the SIR improvement for both the BLCMV-N and the BLCMV. For the BMVDR-N it can be observed that increasing the mixing parameter η reduces the SIR improvement. It should be noted that the exact behaviour depends on the relative position of the interfering source to the desired source, as can be seen from (45).

2) *Binaural Cue Preservation of Background Noise:* For different frequencies, Fig. 6 depicts the input MSC in (27) of the noise component (**Input**) and the output MSC in (27) of the noise component for the BLCMV in (41) for different values of the interference scaling parameter δ , the BMVDR-N for different values of the mixing parameter η and the BLCMV-N for different values of the mixing parameter η and the interference scaling parameter δ . Although the BLCMV is not designed to preserve the MSC of the noise component, it can be observed that an output MSC smaller than 1 is obtained, especially for large values of δ [14]. However, since the output MSC of the noise component depends on the relative position of the interfering source to the desired source, it cannot be easily controlled. Since the BMVDR-N mixes the output signals of the BMVDR with the noisy reference microphone signals, it can be observed that the output MSC of the noise component is smaller than 1, and for $\eta = 1$ the MSC is perfectly preserved (but no beamforming is applied). For the BLCMV-N, it can be observed that both η and δ influence the output MSC of the noise component, as discussed in Section V-C.

For $\eta = 0$, the output MSC of the noise component for the BLCMV-N is obviously equal to the output MSC of the noise component for the BLCMV. For a fixed value of δ , it can be observed that the output MSC of the noise component approaches the input MSC of the noise component for increasing η , although it should be realized that perfect preservation of the MSC of the noise component is only possible for $\delta = 1$ (cf. Section V-C). For several values of the mixing parameter η , Fig. 7 depicts the MSC error of the noise component for the BLCMV-N and the

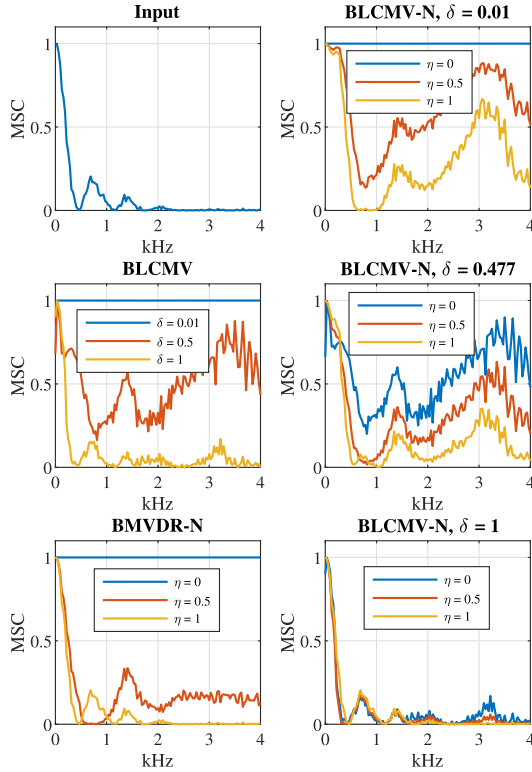


Fig. 6. The MSC of the noise component in the reference microphone signals (**Input**), in the output signals of the BLCMV for different values of the interference scaling parameter δ , the BMVDR-N for different values of the mixing parameter η and the BLCMV-N for different values of the mixing parameter η and the interference scaling parameter δ .

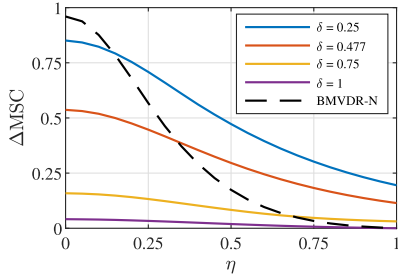


Fig. 7. Frequency-averaged MSC error of the noise component for the BLCMV-N and the BMVDR-N.

BMVDR-N, averaged over all frequencies, i.e.,

$$\Delta \text{MSC} = \frac{1}{F-1} \sum_{f=1}^{F-1} |\text{MSC}_n^{\text{in}}(f) - \text{MSC}_n^{\text{out}}(f)|, \quad (83)$$

with f the frequency bin index and F the total number of frequency bins. As expected, the BMVDR ($\eta = 0$) yields the largest MSC error of the noise component and increasing the mixing parameter η reduces the frequency-averaged MSC error of the noise component for the BMVDR-N [16]. For the considered acoustic scenario, it can be observed for the BLCMV-N that for any value of the interference scaling parameter δ , increasing the mixing parameter η reduces the frequency-averaged MSC error of the noise component compared to the BLCMV ($\eta = 0$). Further, it can be observed that for small values of the interference

scaling parameter δ , the effect of the mixing parameter η is larger than for large values of the interference scaling parameter δ , for which the frequency-averaged MSC error is relatively small for all values of the mixing parameter η . These results clearly show that the mixing parameter η in the BLCMV-N enables to control the binaural cues of the background noise.

B. Experimental Results Using Reverberant Recordings

For a more realistic evaluation, we compare the performance of the considered binaural beamforming algorithms using reverberant recordings. Similarly to Section VI-A, the experimental setup consists of two hearing aids, each with two microphones, mounted on a HATS in a cafeteria with a reverberation time of approximately 1.25 s [24]. The desired source was again placed at 0° (at a distance of about 102 cm), while the interfering source was again placed at -35° (at a distance of about 118 cm), see [24] for more details. The desired and interfering source components were generated by convolving clean speech signals with the measured reverberant room impulse responses corresponding to the desired source and interfering source positions. The desired source was a male German speaker, speaking eight sentences with a pause of 1 s between the sentences. The interfering source was a male Dutch speaker, speaking seven sentences with a pause of 0.25 s between the sentences. As background noise we used realistic recordings [24], consisting of multi-talker babble noise, clacking plates and temporally dominant competing speakers. The used background noise hence clearly differed from the perfectly diffuse noise in Section VI-A. The entire signal had a length of about 28 s. The desired source and the background noise were active the entire time, whereas the interfering source only became active after about 14s. The desired source component, the interfering source component and the noise component were mixed at an input SNR of 10 dB and input SIR of 5 dB in the right reference microphone. Again, we chose the front microphone on each hearing aid as reference microphone.

The processing was performed at a sampling rate of 16 kHz in the STFT domain with a frame length of 8192 samples and a square-root Hann window with 50% overlap. We used an oracle voice activity detector (i.e., using the desired source and interfering source signals) to estimate the noise covariance matrix \mathbf{R}_n , the undesired covariance matrix \mathbf{R}_v (interfering source plus background noise) and $\mathbf{R}_{xn} = \mathbf{R}_x + \mathbf{R}_n$ (desired source plus background noise) over the entire signal. All binaural beamforming algorithms were implemented using relative transfer function (RTF) vectors [25], relating the ATF vectors in (4) to the reference microphones. Using the covariance whitening method (see [14], [26] for further details) the RTF vectors of the desired source and the interfering source were estimated based on generalized eigenvalue decomposition of \mathbf{R}_{xn} and \mathbf{R}_n or \mathbf{R}_v and \mathbf{R}_n , respectively. The mixing parameter was set to $\eta = 0.3$ and the interference scaling parameter was set to $\delta = 0.3$.

As objective performance measures for noise and interference reduction performance, we used the left and the right SNR improvement (ΔSNR_L , ΔSNR_R) and the left and the right SIR improvement (ΔSIR_L , ΔSIR_R). As objective performance measure for binaural cue preservation of the background noise

TABLE II
OBJECTIVE PERFORMANCE MEASURES FOR ALL CONSIDERED ALGORITHMS
IN THE REVERBERANT ENVIRONMENT

	BMVDR	BLCMV	BMVDR-N	BLCMV-N
ΔSNR_L [dB]	13.0	10.1	8.6	7.6
ΔSNR_R [dB]	12.9	9.2	8.6	7.0
ΔSIR_L [dB]	-0.1	9.7	0.82	9.8
ΔSIR_R [dB]	-4.3	8.7	-2.4	8.9
ΔMSC	0.86	0.64	0.10	0.19

we used the frequency-averaged MSC error of the noise component (ΔMSC) as defined in (83). All objective performance measures were computed using the reference microphone signals and the output signals of all considered algorithms. Table II presents the objective performance measures for all considered algorithms.

1) *Noise and Interference Reduction Performance*: In terms of noise reduction performance, it can be observed that – as expected – the BMVDR yields the highest SNR improvement (13.0 dB for the left and 12.9 dB for the right side). All other algorithms yield a lower SNR improvement, for the BLCMV due to the additional constraint for the interfering source, for the BMVDR-N due to the mixing with the noisy reference microphone signals, and for the BLCMV-N due to both effects. The partial noise estimation for the BLCMV-N seems to result in a smaller drop in noise reduction performance compared to the BLCMV (2.5 dB for the left side, 2.2 dB for the right side) than for the BMVDR-N compared to the BMVDR (4.4 dB for the left side, 4.3 dB for the right side). Please note that both for the BMVDR-N as well as for the BLCMV-N this drop in noise reduction performance depends on the relative position of the interfering source to the desired source.

In terms of interference reduction performance, it can be observed that both the BLCMV and the BLCMV-N approximately lead to the same SIR improvement (for the left and the right side), which is in line with the theoretical SIR improvement in (40) and (71), i.e., $10 \log_{10} \frac{1}{\beta^2} \approx 10.5$ dB. The fact that this theoretical SIR improvement is not reached and the fact that the SIR improvements for the BLCMV and BLCMV-N are not exactly the same is due to estimation errors in the covariance matrices, which was also already noted in [14], [17]. In addition, it can be observed that the BMVDR and BMVDR-N lead to very low (even negative) SIR improvements, which is presumably due to the fact that the interfering source is relatively close to the desired source.

2) *Binaural Cue Preservation of Background Noise*: As expected, the BMVDR yields the largest MSC error of the noise component ΔMSC . As discussed in Section III-B, the output MSC of the noise component for the BLCMV is typically smaller than 1, hence leading to a smaller MSC error compared to the BMVDR. Due to the mixing with the noisy reference microphone signals, both the BMVDR-N and the BLCMV-N yield a much smaller MSC error of the noise component than the BMVDR and the BLCMV, where the MSC error is slightly smaller for the BMVDR-N than for the BLCMV-N.

In conclusion, the objective performance measures show that the BLCMV-N leads to a very similar interference reduction as the BLCMV, while providing a trade-off between noise

reduction performance (slightly worse than the BLCMV) and binaural cue preservation of the background noise (much better than the BLCMV). However, it should be noted that compared to the BMVDR-N one of the apparent weaknesses of the proposed BLCMV-N (and BLCMV) is that spatial information of both the desired source and the interfering source is required, which is not straightforward to estimate simultaneously in practice, especially for moving sources.

C. Perceptual Listening Test

To further investigate the spatial impression of the different output signal components for the four considered algorithms, we conducted a perceptual listening test similarly to [21]. The desired source was now placed at -35° and the interfering source was placed at 90° , in order to enhance the perceived spatial differences between both sources. The desired source component, the interfering source component and the noise component were mixed at an input SNR of 0 dB and input SIR of 0 dB in the right reference microphone. Thirteen self-reported normal-hearing subjects participated in the perceptual listening test, where none of the authors participated. All subjects can be considered expert listeners, i.e., they were familiar with similar perceptual listening tests, and gave informed consent. The listening test was conducted in a sound proof listening booth using an RME Fireface UCX sound card with Sennheiser HD 580 headphones.

Using a procedure similar to the Multi-Stimulus Test with Hidden Reference and Anchor (MUSHRA) [27], the task was to rate the perceived spatial difference with respect to a reference signal. For a coherent source (e.g., interfering source), this corresponds to rating differences in perceived source location, whereas for a diffuse noise field this corresponds to rating differences in perceived diffuseness. A score of 0 is associated with a large perceived spatial difference, whereas a score of 100 is associated with no perceived spatial difference. As reference signal we used the (unprocessed) reference microphone signals, while as anchor signal we used the left reference microphone signal, played back to both ears. The anchor signal was hence a monaural signal with no binaural cues, which is perceived in the center of the head.

We conducted three evaluations, where only some components were active in the output signals, the reference signal and the anchor signal. In the first evaluation, only the desired source component and the interfering source component (i.e., no noise component) were active and the task was to rate the spatial difference for the interfering source. In the second evaluation, only the desired source component and the noise component (i.e., no interfering source component) were active and the task was to rate the spatial difference for the background noise. In the third evaluation, all signal components were active and the task was to rate the spatial difference for the interfering source and the background noise simultaneously. To familiarize the subjects with the tasks and the sound material, a training round was performed. Audio samples for all binaural beamforming algorithms and the unprocessed input signals are

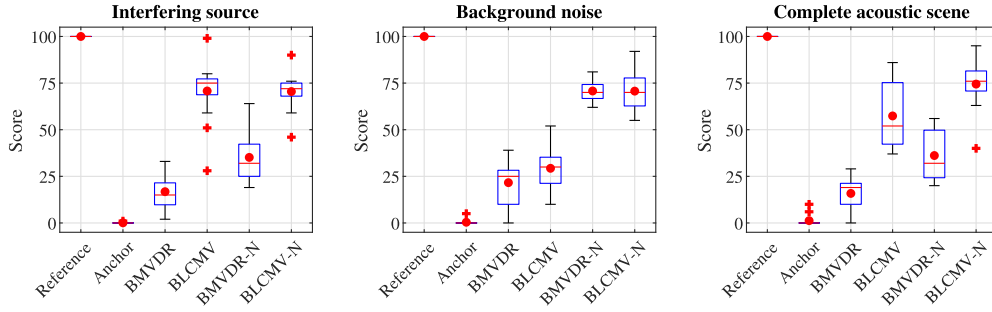


Fig. 8. Boxplot of the MUSHRA scores for all three evaluations. The plot depicts the median score (red line), the mean score (red dot), the first and third quartiles (blue boxes) and the interquartile ranges (whiskers). Outliers are indicated by red + markers.

available online (see <https://uol.de/en/sigproc/research/audio-demos/binaural-noise-reduction/blcmv-n-beamformer>).

The MUSHRA scores for the three evaluations are shown in Fig. 8. A one-way repeated-measures ANOVA was performed. The analysis revealed a significant within-subjects effect for all three evaluations. Hence, post-hoc comparison t-tests with Bonferroni correction were performed [28].

a) Interfering source: The within-subjects effect was significant [$F(2.098, 25.176) = 219.2$, $p < .001$, Greenhouse-Geisser correction]. As expected, the BLCMV and the BLCMV-N preserved the spatial impression of the interfering source significantly better than the BMVDR and the BMVDR-N ($p < .001$). The BMVDR-N performed significantly better than the BMVDR ($p < .001$), which is not unexpected since the interfering source component is also mixed with the mixing parameter η . No significant difference was found between the BLCMV and the BLCMV-N ($p = 1$).

b) Background noise: The within-subjects effect was significant [$F(3.072, 36.869) = 332.066$, $p < .001$, Greenhouse-Geisser correction]. As expected, the BMVDR-N and the BLCMV-N, both using partial noise estimation, preserved the spatial impression of the background noise significantly better than the BMVDR and the BLCMV ($p < .001$). No significant difference was found between the BMVDR-N and the BLCMV-N ($p = 1$) and between the BMVDR and BLCMV ($p = .614$).

c) Complete acoustic scene: The within-subjects effect was significant [$F(2.905, 34.858) = 171.783$, $p < .001$, Greenhouse-Geisser correction]. In terms of preservation of the spatial impression of the complete acoustic scene, the BMVDR-N scored significantly higher than the BMVDR ($p < .001$), the BLCMV scored significantly higher than the BMVDR-N ($p = .014$), and the proposed BLCMV-N scored significantly higher than the BLCMV ($p = .025$).

In summary, the results of the listening test showed that the BLCMV-N is capable of preserving the spatial impression of an interfering source and background noise in a realistic acoustic scenario, outperforming all other considered binaural beamforming algorithms in terms of spatial impression. In future work we plan to perform a subjective evaluation using hearing-impaired subjects.

VII. CONCLUSION

In this paper we proposed the BLCMV-N, merging the advantages of the BLCMV and the BMVDR-N, i.e., preserving

the binaural cues of the interfering source and controlling the reduction of the interfering source as well as the binaural cues of the background noise. We showed that the output signals of the BLCMV-N can be interpreted as a mixture between the noisy reference microphone signals and the output signals of a BLCMV using an adjusted interference scaling parameter. We provided a theoretical comparison between the BMVDR, the BLCMV, the BMVDR-N and the proposed BLCMV-N in terms of noise and interference reduction performance and binaural cue preservation. The obtained analytical expressions were first validated using measured anechoic acoustic transfer functions. Experimental results using recorded signals in a realistic reverberant environment showed that the BLCMV-N leads to a very similar interference reduction as the BLCMV, while providing a trade-off between noise reduction performance (slightly worse than the BLCMV) and binaural cue preservation of the background noise (much better than the BLCMV). In addition, the results of a perceptual listening test with 13 normal-hearing participants showed that the proposed BLCMV-N is capable of preserving the spatial impression of an interfering source and background noise in a realistic acoustic scenario, outperforming all other considered binaural beamforming algorithms in terms of spatial impression.

APPENDIX A

DERIVATION OF THE BLCMV-N

Using (4), (6) and (36), the constrained optimization problem in (46) can be reformulated as

$$\min_{\mathbf{w}_L} \mathcal{E} \left\{ \left| \mathbf{w}_L^H \mathbf{n} - \eta m_L \right|^2 \right\} \quad \text{s.t.} \quad \mathbf{C}^H \mathbf{w}_L = \mathbf{g}_L. \quad (84)$$

This constrained optimization problem can be solved using the method of Lagrange multipliers, where the Lagrangian function is given by

$$\begin{aligned} \mathcal{L}(\mathbf{w}_L, \boldsymbol{\lambda}_L) &= \mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_L - \eta \boldsymbol{\eta}^T \mathbf{R}_n \mathbf{w}_L - \eta \mathbf{w}_L^H \mathbf{R}_n \mathbf{e}_L \\ &+ \eta^2 p_{n,L}^{\text{in}} + \boldsymbol{\lambda}_L^H (\mathbf{C}^H \mathbf{w}_L - \mathbf{g}_L) + (\mathbf{w}_L^H \mathbf{C} - \mathbf{g}_L^H) \boldsymbol{\lambda}_L, \end{aligned} \quad (85)$$

with $\boldsymbol{\lambda}_L$ denoting the 2-dimensional vector of Lagrangian multipliers. Setting the gradient with respect to \mathbf{w}_L

$$\nabla_{\mathbf{w}_L} \mathcal{L}(\mathbf{w}_L, \boldsymbol{\lambda}_L) = 2\mathbf{R}_n \mathbf{w}_L - 2\eta \mathbf{R}_n \mathbf{e}_L + 2\mathbf{C} \boldsymbol{\lambda}_L \quad (86)$$

equal to $\mathbf{0}$ yields

$$\mathbf{w}_L = \eta \mathbf{e}_L - \mathbf{R}_n^{-1} \mathbf{C} \boldsymbol{\lambda}_L. \quad (87)$$

Substituting (87) into the constraint $\mathbf{C}^H \mathbf{w}_L = \mathbf{g}_L$ and solving for the Lagrangian multiplier λ_L yields

$$\lambda_L = (\mathbf{C}^H \mathbf{R}_n^{-1} \mathbf{C})^{-1} (\eta \mathbf{C}^H \mathbf{e}_L - \mathbf{g}_L). \quad (88)$$

Substituting (88) into (87), the solution to (46) is given by

$$\begin{aligned} \mathbf{w}_{\text{BLCMV-N},L} = \\ \eta \mathbf{e}_L + \mathbf{R}_n^{-1} \mathbf{C} (\mathbf{C}^H \mathbf{R}_n^{-1} \mathbf{C})^{-1} (\mathbf{g}_L - \eta \mathbf{C}^H \mathbf{e}_L), \end{aligned} \quad (89)$$

where, using (36),

$$\mathbf{g}_L - \eta \mathbf{C}^H \mathbf{e}_L = \begin{bmatrix} (1 - \eta) a_L^* \\ (\delta - \eta) b_L^* \end{bmatrix}. \quad (90)$$

APPENDIX B

OUTPUT NOISE PSD FOR THE BLCMV-N

Using (59) in (16) with \mathbf{R}_n instead of \mathbf{R}_x , the output PSD of the noise component for the BLCMV-N is given by

$$\begin{aligned} \mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_L &= \eta^2 \mathbf{e}_L^T \mathbf{R}_n \mathbf{e}_L \\ &+ \eta(1 - \eta) [a_L \mathbf{w}_x^H \mathbf{R}_n \mathbf{e}_L + \mathbf{e}_L^T \mathbf{R}_n \mathbf{w}_x a_L^*] \\ &+ \eta(\delta - \eta) [b_L \mathbf{w}_u^H \mathbf{R}_n \mathbf{e}_L + \mathbf{e}_L^T \mathbf{R}_n \mathbf{w}_u b_L^*] \\ &+ (1 - \eta)^2 |a_L|^2 \mathbf{w}_x^H \mathbf{R}_n \mathbf{w}_x \\ &+ (\delta - \eta)(1 - \eta) [a_L^* b_L \mathbf{w}_u^H \mathbf{R}_n \mathbf{w}_x + a_L b_L^* \mathbf{w}_x^H \mathbf{R}_n \mathbf{w}_u] \\ &+ (\delta - \eta)^2 |b_L|^2 \mathbf{w}_u^H \mathbf{R}_n \mathbf{w}_u. \end{aligned} \quad (91)$$

Using (56) and (58), the components in (91) are given by [14]

$$\begin{aligned} \mathbf{e}_L^T \mathbf{R}_n \mathbf{w}_x &= \frac{1}{1 - \Psi} \left(\frac{a_L}{\gamma_a} - \Psi \frac{b_L}{\gamma_{ab}} \right), \\ \mathbf{w}_x^H \mathbf{R}_n \mathbf{w}_x &= \frac{1}{(1 - \Psi) \gamma_a}, \\ \mathbf{e}_L^T \mathbf{R}_n \mathbf{w}_u &= \frac{1}{1 - \Psi} \left(\frac{b_L}{\gamma_b} - \Psi \frac{a_L}{\gamma_{ab}^*} \right), \\ \mathbf{w}_u^H \mathbf{R}_n \mathbf{w}_u &= \frac{1}{(1 - \Psi) \gamma_b}, \\ \mathbf{w}_x^H \mathbf{R}_n \mathbf{w}_u &= \frac{\Psi}{(1 - \Psi) \gamma_{ab}^*}. \end{aligned} \quad (92)$$

Substituting (92) in (91) yields

$$\begin{aligned} \mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_L &= \eta^2 p_{n,L}^{\text{in}} + \frac{1}{1 - \Psi} \left[(1 - \eta^2) \frac{|a_L|^2}{\gamma_a} + (\delta^2 - \eta^2) \frac{|b_L|^2}{\gamma_b} \right. \\ &\left. - 2\Psi(\delta - \eta^2) \Re \left\{ \frac{a_L b_L^*}{\gamma_{ab}^*} \right\} \right] = \mathbf{e}_L^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{xu,3}) \mathbf{e}_L, \end{aligned} \quad (93)$$

with $\mathbf{R}_{xu,3}$ defined in (66). Similarly, it can be shown that

$$\mathbf{w}_L^H \mathbf{R}_n \mathbf{w}_R = \mathbf{e}_L^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{xu,3}) \mathbf{e}_R, \quad (94)$$

$$\mathbf{w}_R^H \mathbf{R}_n \mathbf{w}_R = \mathbf{e}_R^T (\eta^2 \mathbf{R}_n + \mathbf{R}_{xu,3}) \mathbf{e}_R. \quad (95)$$

REFERENCES

- [1] V. Hamacher, U. Kornagel, T. Lotter, and H. Puder, "Binaural signal processing in hearing aids: Technologies and algorithms," in *Adv. in Digit. Speech Transmiss.* New York, NY, USA: Wiley, 2008, pp. 401–429.
- [2] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.
- [3] S. Doclo, S. Gannot, D. Marquardt, and E. Hadad, "Binaural speech processing with application to hearing devices," in *Audio Source Separation and Speech Enhancement*. Berlin, Germany: Wiley, 2018, ch. 18, pp. 413–442.
- [4] J. Blauert, *Spatial Hear.: The Psychophysics of Hum. Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.
- [5] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality," *J. Acoust. Soc. Amer.*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.
- [6] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Amer.*, vol. 83, no. 4, pp. 1508–1516, Apr. 1988.
- [7] M. L. Hawley, R. Y. Litovsky, and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Amer.*, vol. 115, no. 2, pp. 833–843, Feb. 2004.
- [8] D. P. Welker, J. E. Greenberg, J. G. Desloge, and P. M. Zurek, "Microphone-array hearing aids with binaural output. II. A two-microphone adaptive system," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 543–551, Nov. 1997.
- [9] R. Aichner, H. Buchner, M. Zourub, and W. Kellermann, "Multi-channel source separation preserving spatial information," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2007, pp. 5–8.
- [10] T. Klases, T. van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.
- [11] B. Cornelis, S. Doclo, T. van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural multi-microphone noise reduction techniques," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.
- [12] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function MVDR beamformers with interference cue preservation constraints," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.
- [13] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 12, pp. 2162–2176, Dec. 2015.
- [14] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 3, pp. 543–558, Mar. 2016.
- [15] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, "Relaxed binaural LCMV beamforming," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 1, pp. 137–152, Jan. 2017.
- [16] D. Marquardt and S. Doclo, "Interaural coherence preservation for binaural noise reduction using partial noise estimation and spectral postfiltering," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 7, pp. 1257–1270, Jan. 2018.
- [17] N. Göbbling, D. Marquardt, I. Merks, T. Zhang, and S. Doclo, "Optimal binaural LCMV beamforming in complex acoustic scenarios: Theoretical and practical insights," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Sep. 2018, pp. 381–385.
- [18] H. As'ad, M. Bouchard, and H. Kamkar-Parsi, "A robust target linearly constrained minimum variance beamformer with spatial cues preservation for binaural hearing aids," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 10, pp. 1549–1563, Oct. 2019.
- [19] R. M. Corey and A. C. Singer, "Binaural audio source remixing with microphone array listening devices," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2020, pp. 561–565.
- [20] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Comparison of binaural multichannel Wiener filters with binaural cue preservation of the interferer," in *Proc. IEEE Int. Conf. Sci. Elect. Eng.*, Nov. 2016, pp. 1–5.
- [21] N. Göbbling, D. Marquardt, and S. Doclo, "Perceptual evaluation of binaural MVDR-based algorithms to preserve the interaural coherence of diffuse noise fields," *Trends Hear.*, vol. 24, pp. 1–18, Apr. 2020.
- [22] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

- [23] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 4, pp. 692–730, Apr. 2017.
- [24] H. Kayser, S. D. Ewert, J. Anemüller, 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.*, vol. 2009, pp. 1–10, Jan. 2009.
- [25] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and non-stationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [26] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.
- [27] *Method for the Subjective Assessment of Intermediate Qual. Level of Coding Syst.*, ITU-R BS.1534-1, International Telecommunications Union (ITU-T) Recommendation, Geneva, Switzerland, Jan. 2003.
- [28] B. R. Kirkwood and J. A. C. Sterne, *Essential Medical Statistics*. Hoboken, NJ, USA: Wiley, 2010.



Nico Gößling received the B.Sc. degree in media technology with focus on audio technology from the Hamburg University of Applied Sciences, Hamburg, Germany, in 2013 and the M.Sc. degree in hearing technology and audiology from the University of Oldenburg, Oldenburg, Germany, in 2016. From 2016 to 2020, he was a Doctoral Researcher with the Signal Processing Division, Department of Medical Physics and Acoustics, University of Oldenburg. His research interests are in the area of signal processing for acoustical applications, more specifically microphone array

processing for binaural hearing devices and acoustic sensor networks. In 2018, he received the Best Student Paper Award at the 16th International Workshop on Acoustic Signal Enhancement (IWAENC) in Tokyo, Japan. He is currently a Patent Attorney Candidate in Hamburg, Germany.



Elijor Hadad received the B.Sc. (*summa cum laude*) and M.Sc. (*cum laude*) degrees in electrical engineering from the Ben-Gurion University of the Negev (BGU), Beer-Sheva, Israel, in 2001 and 2007, respectively, and the Ph.D. degree from Bar-Ilan University (BIU), Ramat Gan, Israel, in 2017. He is a Postdoctoral Researcher with the Audio Processing Laboratory, Faculty of Engineering, Bar-Ilan University. His research interests include statistical signal processing and in particular binaural noise reduction algorithms using microphone arrays, and speaker localization

and tracking.



Sharon Gannot (Senior Member, IEEE) received the 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, Tel Aviv, Israel, in 1995 and 2000, respectively, all in electrical engineering.

In 2001, he held a Postdoctoral position with the Department of Electrical Engineering, KU Leuven, Leuven, Belgium. From 2002 to 2003, he held a Research and Teaching position with the Faculty of Electrical Engineering, Technion-Israel Institute of

Technology. He is currently a Full Professor with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel, where he is heading the Acoustic Signal Processing Laboratory and the Data Science Program. He also serves as the Faculty Vice Dean and as the Bar-Ilan University Data Science Institute Deputy Director. In 2018–2019, he was also a part-time Professor with the Technical Faculty of IT and Design, Aalborg University, Denmark. His research interests include statistical signal processing and machine learning algorithms (including manifold learning and deep learning) with applications to single- and multi-microphone speech processing. Specifically, distributed algorithms for ad hoc microphone arrays, speech enhancement, noise reduction and speaker separation, dereverberation, single microphone speech enhancement, speaker localization, and tracking.

Dr. Gannot was an Associate Editor for the *EURASIP Journal of Advances in Signal Processing* during 2003–2012, and an Editor for several special issues on multi-microphone speech processing of the same journal. He was a Guest Editor for several special issues for Elsevier's *Speech Communication and Signal Processing* journals. He was also the Lead Guest Editor of a special issue on speaker localization for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, 2019. He was an Associate Editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING in 2009–2013, and the Area Chair for the same journal in 2013–2017. He is currently a Moderator for arXiv in the field of audio and speech processing. He is also a Reviewer for many IEEE journals and conferences. Since January 2010, he has been a Member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE Signal Processing Society (SPS) and served as the Committee Chair, 2017–2018. He is currently a member of the Data Science Initiative of IEEE SPS. He is also a member of the IEEE SPS, Conferences Board since 2019 and a member of the Executive Subcommittee of this board since 2020. Since 2020, he is also a member of the IEEE SPS Education Board. Since 2005, he has been a member of the technical and steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) and was the General Co-Chair of the IWAENC held in Tel-Aviv, Israel, in August 2010. He was the General Co-Chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in October 2013. He was selected (with colleagues) to present tutorial sessions at ICASSP 2012, EUSIPCO 2012, ICASSP 2013, EUSIPCO 2013 and EUSIPCO 2019, and was a keynote speaker for IWAENC 2012, International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA) 2017, Informationstechnische Gesellschaft im VDE (ITG) Conference on Speech Communication 2018 and Audio Analysis Workshop 2018. He was the recipient of Bar-Ilan University Outstanding Lecturer Award in 2010 and 2014 and the Rector Innovation in Research Award in 2018. He was also the co-recipient of twelve best paper awards.



Simon Doclo (Senior Member, IEEE) 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, respectively. From 2003 to 2007, he was a Postdoctoral Fellow with the Research Foundation Flanders, Electrical Engineering Department (Katholieke Universiteit Leuven) and the Cognitive Systems Laboratory (McMaster University, Canada). From 2007 to 2009, he was a Principal Scientist with NXP Semiconductors, Leuven, Belgium. Since 2009, he is a Full Professor

with the University of Oldenburg, Germany, and scientific advisor for the Division Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research activities center around signal processing for acoustical and biomedical applications, more specifically microphone array processing, speech enhancement, active noise control, acoustic sensor networks and hearing aid processing. He received several best paper awards (International Workshop on Acoustic Echo and Noise Control 2001, EURASIP Signal Processing 2003, IEEE Signal Processing Society 2008, VDE Information Technology Society 2019). He is member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, the EURASIP Technical Area Committee on Acoustic, Speech and Music Signal Processing and the EAA Technical Committee on Audio Signal Processing. He was and is involved in several large-scale national and European research projects (ITN DREAMS, Cluster of Excellence Hearing4all, CRC Hearing Acoustics). He was Technical Program Chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in 2013 and Chair of the ITG Conference on Speech Communication in 2018. In addition, he served as a Guest Editor for several special issues (*IEEE Signal Processing Magazine*, Elsevier's *Signal Processing*) and was an Associate Editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and *EURASIP Journal on Advances in Signal Processing*.