Performance Analysis of the Extended Binaural MVDR Beamformer With Partial Noise Estimation

Nico Gößling, Daniel Marquardt, *Member, IEEE*, and Simon Doclo, *Senior Member, IEEE*

***Abstract*—Besides reducing undesired noise sources and limiting speech distortion, another important objective of a binaural noise reduction algorithm is the preservation of the binaural cues of all sound sources in the acoustic scene. In this paper, we consider the binaural minimum variance distortionless response beamformer with partial noise estimation (BMVDR-N), which allows to trade off between noise reduction performance and binaural cue preservation of the noise component by mixing the output signals of the BMVDR beamformer with the noisy reference microphone signals. For a directional noise source, it has been shown that incorporating an external microphone in addition to the head-mounted microphones enables both the noise reduction performance as well as the interaural time and level difference cues of the noise component to be improved in the output signals. In this paper, we consider an arbitrary noise field and analytically show that incorporating an external microphone in the BMVDR-N beamformer enables 1) a larger output signal-to-noise ratio (SNR) for the same mixing parameter, 2) the same output SNR for a larger mixing parameter, and 3) the same desired output magnitude squared coherence (MSC) of the noise component for a smaller mixing parameter to be obtained. The derived analytical expressions are firstly validated using simulated anechoic acoustic transfer functions, where the listener's head is modelled as a rigid sphere. Experimental results using recorded signals for a binaural hearing device setup in a reverberant environment also show that in a realistic scenario incorporating an external microphone in the BMVDR-N beamformer significantly improves the output SNR and reduces the mixing parameter that is required to obtain a desired output MSC of the noise component compared to using only the head-mounted microphones.**

***Index Terms*—Binaural cues, binaural noise reduction, external microphone, hearing devices, speech enhancement.**

I. Introduction

NOISE reduction algorithms for head-mounted assistive hearing devices (e.g., hearing aids, earbuds or hearables) are crucial to improve speech intelligibility and speech quality in noisy environments. Compared to a bilateral configuration where both devices operate independently, in a binaural configuration both devices exchange information, such that the information captured by all microphones on both sides of the head can be exploited [1]–[3]. Besides reducing background noise and limiting speech distortion, another important objective of a binaural noise reduction algorithm is the preservation of the listener's spatial perception of the acoustic scene, in order to exploit the binaural hearing advantage [4]–[6] and to reduce confusions due to a possible mismatch between acoustical and visual information. This can be achieved by preserving the binaural cues of all sound sources, i.e., the desired source as well as the undesired noise. For a coherent (directional) source the main binaural cues used by the auditory system are the interaural level difference (ILD) and the interaural time difference (ITD) [5]. For incoherent sources (e.g., a diffuse noise field) the ILD and ITD are not very descriptive, but the interaural coherence (IC) and the magnitude-squared coherence (MSC) are known to play a major role in spatial perception, e.g., to determine the spatial width or diffuseness [5], [7]. In practice, diffuse-like noise fields for example arise when multiple people are talking around the listener, i.e. the well-known "cocktail-party scenario" [8].

For a single desired source, it has been shown that the binaural multi-channel Wiener filter (MWF) and the binaural minimum variance distortionless response (BMVDR) beamformer, as a special case of the binaural MWF, preserve the binaural cues of the desired source component but distort the binaural cues of the noise component [2], [3], [9]. More precisely, after applying the BMVDR beamformer both output components exhibit the binaural cues of the desired source component, i.e., both components are perceived as coming from the direction of the desired source such that the binaural hearing advantage cannot be exploited by the auditory system. Aiming at additionally preserving the binaural cues of the noise component and hence preserving the spatial impression of the complete acoustic scene, several extensions of the binaural MWF and the BMVDR beamformer have been proposed, e.g., by incorporating constraints into the spatial filter design [10]–[15] or by mixing with scaled (noisy) reference microphone signals [9], [16]–[20].

In this paper we focus on the BMVDR beamformer with partial noise estimation (BMVDR-N) [17], [18], which mixes the output signals of the BMVDR beamformer with the noisy reference microphone signals using a mixing parameter. The mixing parameter allows to trade off between noise reduction performance and binaural cue preservation of the noise component. In order to achieve a desired output MSC of the noise

Manuscript received March 6, 2020; revised August 27, 2020 and October 19, 2020; accepted November 29, 2020. Date of publication December 10, 2020; date of current version December 30, 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 joint Lower Saxony-Israeli Project ATHENA. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jun Du. *(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@uni-oldenburg.de; simon.doclo@uni-oldenburg.de).

Daniel Marquardt was with the University of Oldenburg, 26111 Oldenburg, Germany. He is now with Starkey Hearing Technologies, Eden Prairie, MN 55344 USA (e-mail: daniel_marquardt@starkey.de).

Digital Object Identifier 10.1109/TASLP.2020.3043674

component, in [18] a closed-form expression for the mixing parameter of the BMVDR-N beamformer has been derived. Using the mixing parameter calculated with this closed-form expression in the BMVDR-N beamformer therefore leads to a pre-defined binaural cue preservation of the noise component. Setting the desired output MSC of the noise component close to the input MSC of the noise component leads to a better binaural cue preservation, but typically to a lower noise reduction performance. The desired output MSC of the noise component can be psycho-acoustically motivated based on the IC discrimination ability of the human auditory system [18], [21], aiming for the spatial impression of the noise component in the reference microphone signals and the noise component in the output signals to be indistinguishable.

To improve the performance of (binaural) noise reduction algorithms, it has been proposed to use one or more external microphones (e.g., lying on a table, attached to a person) in conjunction with the head-mounted microphones [22]–[30]. Such external microphones make it possible to not only locally sample the sound field (at the listener's head) but increase spatial diversity by spatially distributing the microphones. For a coherent (directional) noise source, it has been shown in [23] that incorporating an external microphone in the binaural MWF with partial noise estimation enables both the output signal-to-noise ratio (SNR) as well as the binaural cues, i.e., ILD and ITD, of the output noise component to be improved compared to only using the head-mounted microphones.

In this paper, we consider an arbitrary noise field and derive analytical expressions for the output SNR and the binaural cues (more in particular the MSC) of the output noise component when incorporating an external microphone in the BMVDR-N beamformer. First, we show that an external microphone either enables a larger output SNR to be obtained for the same mixing parameter or the same output SNR for a larger mixing parameter compared to using only the head-mounted microphones. Secondly, we show that the same desired output MSC of the noise component can be obtained for a smaller mixing parameter, implying that an external microphone enables the same spatial impression of the noise component to be achieved compared to using only the head-mounted microphones while achieving a larger output SNR. Preliminary results were published in [31] for the special case of a homogeneous noise field and assuming the desired source is in front of the listener, whereas in this paper no assumptions are made with respect to the noise field or the position of the desired source. The derived analytical expressions are firstly validated using simulated anechoic acoustic transfer functions (ATFs), where the listener's head is modelled as a rigid sphere [32]. Experiments are then performed using recorded signals for a binaural hearing device configuration in a reverberant environment with multiple interfering talkers as background noise [33]. The experimental results show that in a realistic scenario, incorporating an external microphone in the BMVDR-N beamformer significantly increases the output SNR and decreases the mixing parameter required to obtain a desired output MSC, i.e., the spatial impression, of the noise component for different positions of the external microphone and the desired source.

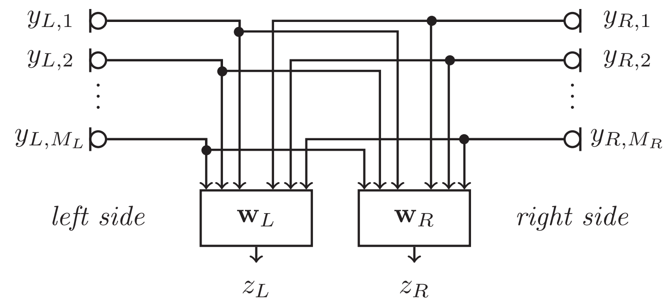


Fig. 1. Binaural hearing device configuration with M_L microphones on the left side and M_R microphones on the right side.

The remainder of this paper is organized as follows. In Section II we introduce the considered binaural hearing device configuration with and without incorporating an external microphone and the used performance measures. In Section III we briefly review two binaural noise reduction algorithms using only the head-mounted microphones: the BMVDR beamformer and the BMVDR-N beamformer. In Section IV we first define the extended BMVDR (eBMVDR) beamformer and the extended BMVDR-N (eBMVDR-N) beamformer, incorporating the external microphone. We then derive analytical expressions for the output SNR and the output MSC of the noise component for the eBMVDR-N beamformer for an arbitrary noise field and without assuming a specific position of the desired source. In Section V we provide simulation results using simulated anechoic ATFs as well as using recorded signals in a reverberant environment.

II. Hearing Device Configurations

In this section, we first consider the binaural hearing device configuration using only head-mounted microphones (Section II-A) and then the extended binaural hearing device configuration, incorporating an external microphone (Section II-B). In Section II-C we define the used performance measures and the binaural cues.

A. Binaural Hearing Device Configuration

Consider the (head-mounted) binaural hearing device configuration depicted in Fig. 1, consisting of M_L microphones on the left side and M_R microphones on the right side, i.e., $M_H = M_L + M_R$. In the frequency-domain, the m-th microphone signal of the left hearing device $y_{L,m}(\omega)$ can be decomposed as

$$y_{L,m}(\omega) = x_{L,m}(\omega) + n_{L,m}(\omega)\,, \quad m \in \{1, \ldots, M_L\}\,, \quad (1)$$

where $x_{L,m}(\omega)$ and $n_{L,m}(\omega)$ denote the desired source component and the noise component in the m-th microphone signal, respectively, and ω denotes the normalized (radian) frequency. Similarly, the m-th microphone signal of the right hearing device can be decomposed as $y_{R,m}(\omega) = x_{R,m}(\omega) + n_{R,m}(\omega)$, $m \in \{1, \ldots, M_R\}$. For conciseness, the variable ω will be omitted in the remainder of the paper. We assume that all head-mounted microphone signals are transmitted (e.g., via a wireless link) between the left and the right hearing device without transmission delay and quantization errors and are synchronized.

The M_H-dimensional stacked input vector containing all microphone signals from both hearing devices is defined as

$$\mathbf{y} = [y_{L,1}, \ldots, y_{L,M_L}, y_{R,1}, \ldots, y_{R,M_R}]^T \in \mathbb{C}^{M_H}, \quad (2)$$

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

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

where $\mathbf{x}$ and $\mathbf{n}$ are defined similarly as $\mathbf{y}$ in (2). In the case of a single desired source, the desired source component can be written as

$$\mathbf{x} = \mathbf{a}s, \quad (4)$$

where s denotes the desired source signal and $\mathbf{a}$ contains the ATFs between the desired source and the head-mounted microphones, including reverberation, microphone characteristics and the head shadow effect.

Without loss of generality, we define the first microphone on both hearing devices as the so-called reference microphones. For conciseness, the reference microphone signals $y_{L,1}$ and $y_{R,1}$ 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 $\mathbf{e}_R$ are M_H-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

$$y_L = x_L + n_L = a_L s + n_L, \quad (6)$$

$$y_R = x_R + n_R = a_R s + n_R. \quad (7)$$

The M_H-dimensional relative transfer function (RTF) vectors of the desired source, relating the ATF vector $\mathbf{a}$ to the reference microphones [34], are defined as

$$\mathbf{a}_L = \frac{\mathbf{a}}{a_L}, \quad \mathbf{a}_R = \frac{\mathbf{a}}{a_R}. \quad (8)$$

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

$$\mathbf{R}_y = \mathcal{E}\left\{\mathbf{y}\mathbf{y}^H\right\}, \ \mathbf{R}_x = \mathcal{E}\left\{\mathbf{x}\mathbf{x}^H\right\}, \ \mathbf{R}_n = \mathcal{E}\left\{\mathbf{n}\mathbf{n}^H\right\}, \quad (9)$$

where $\mathcal{E}\{\cdot\}$ denotes the expected value operator and $(\cdot)^H$ denotes the conjugate transpose. Assuming statistical independence between the desired source and the noise components, $\mathbf{R}_y$ can be written as $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_n$. Using (4) and (9), the desired source covariance matrix can be written as a rank-1 matrix, i.e.,

$$\mathbf{R}_x = p_s \mathbf{a}\mathbf{a}^H, \quad (10)$$

with $p_s = \mathcal{E}\{|s|^2\}$ the power spectral density (PSD) of the desired source signal. The noise covariance matrix $\mathbf{R}_n$ is assumed to be full-rank, i.e., invertible and positive definite.

In the binaural hearing device configuration the output signals z_L and z_R of the left and the right hearing device are obtained by filtering and summing all head-mounted microphone signals (cf. Fig. 1), i.e.,

$$z_L = \mathbf{w}_L^H \mathbf{y}, \quad z_R = \mathbf{w}_R^H \mathbf{y}, \quad (11)$$

where $\mathbf{w}_L$ and $\mathbf{w}_R$ denote the M_H-dimensional (complex-valued) left and right filter vectors, respectively.

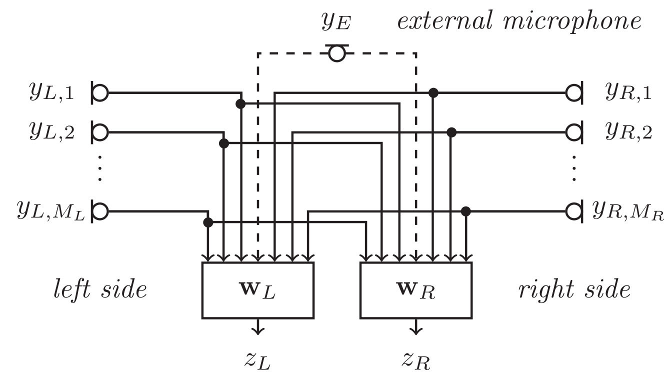


Fig. 2. Extended binaural hearing device configuration with M_L microphones on the left side, M_R microphones on the right side and one external microphone.

B. Extended Binaural Hearing Device Configuration

Fig. 2 depicts the *extended* binaural hearing device configuration, incorporating an additional external microphone. Similarly to (6) and (7), the external microphone signal can be written as

$$y_E = x_E + n_E = a_E s + n_E, \quad (12)$$

where x_E and n_E denote the desired source component and the noise component in the external microphone signal, respectively, and a_E denotes the ATF between the desired source and the external microphone. Similarly as in [23]–[30], we assume that the external microphone signal is transmitted (e.g., via a wireless link) to the binaural hearing devices without transmission delay and quantization errors and that the head-mounted microphone signals and the external microphone signal are synchronized. Methods addressing these problems can be found, e.g., in [35]–[38].

The M-dimensional extended input vector is defined as

$$\mathbf{y}_e = \begin{bmatrix} \mathbf{y} \\ y_E \end{bmatrix} \in \mathbb{C}^M, \quad (13)$$

with $M = M_H + 1$. The extended desired source component $\mathbf{x}_e$, the extended noise component $\mathbf{n}_e$ and the extended ATF vector $\mathbf{a}_e$ are defined similarly as in (13). Similarly to (5), the reference microphone signals can be selected as $y_L = \mathbf{e}_L^T \mathbf{y}_e$ and $y_R = \mathbf{e}_R^T \mathbf{y}_e$, where $\mathbf{e}_L$ and $\mathbf{e}_R$ now are M-dimensional selection vectors. Further, the external microphone signal can be selected as $y_E = \mathbf{e}_E^T \mathbf{y}_e$, where $\mathbf{e}_E$ is an M-dimensional selection vector with $\mathbf{e}_E(M) = 1$. Similarly to (8), the M-dimensional extended RTF vectors of the desired source are defined as

$$\mathbf{a}_{L,e} = \frac{\mathbf{a}_e}{a_L}, \quad \mathbf{a}_{R,e} = \frac{\mathbf{a}_e}{a_R}. \quad (14)$$

Similarly to (9) and (10), the extended desired source and noise covariance matrices are given by

$$\mathbf{R}_{x,e} = \mathcal{E}\left\{\mathbf{x}_e\mathbf{x}_e^H\right\} = p_s\mathbf{a}_e\mathbf{a}_e^H\,, \tag{15}$$

$$\mathbf{R}_{n,e} = \mathcal{E}\left\{\mathbf{n}_e\mathbf{n}_e^H\right\} = \left[\begin{array}{c|c}\mathbf{R}_n & \mathbf{r}_{n,E}\\ \mathbf{r}_{n,E}^H & p_{n_E}\end{array}\right]\,, \tag{16}$$

with

$$\mathbf{r}_{n,E} = \mathcal{E}\left\{\mathbf{n}n_E^*\right\}\,, \tag{17}$$

the cross-correlation vector between the noise component in the head-mounted microphone signals and the noise component in the external microphone signal, and $p_{n_E} = \mathcal{E}\{|n_E|^2\}$ the PSD of the noise component in the external microphone signal. $(\cdot)^*$ denotes complex conjugation. The extended noisy input covariance matrix is given by $\mathbf{R}_{y,e} = \mathbf{R}_{x,e} + \mathbf{R}_{n,e}$.

Similarly to (11), for the extended binaural hearing device configuration the output signals z_L and z_R are obtained by filtering and summing all microphone signals, i.e., the head-mounted microphone signals and the external microphone signal, i.e.,

$$z_L = \mathbf{w}_L^H\mathbf{y}_e\,,\quad z_R = \mathbf{w}_R^H\mathbf{y}_e\,, \tag{18}$$

where $\mathbf{w}_L$ and $\mathbf{w}_R$ now are M-dimensional filter vectors.

C. Performance Measures and Binaural Cues

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

$$\mathrm{SNR}_L^{\mathrm{in}} = \frac{\mathcal{E}\left\{|x_L|^2\right\}}{\mathcal{E}\left\{|n_L|^2\right\}} = \frac{\mathbf{e}_L^T\mathbf{R}_{x,e}\mathbf{e}_L}{\mathbf{e}_L^T\mathbf{R}_{n,e}\mathbf{e}_L}\,, \tag{19}$$

$$\mathrm{SNR}_R^{\mathrm{in}} = \frac{\mathcal{E}\left\{|x_R|^2\right\}}{\mathcal{E}\left\{|n_R|^2\right\}} = \frac{\mathbf{e}_R^T\mathbf{R}_{x,e}\mathbf{e}_R}{\mathbf{e}_R^T\mathbf{R}_{n,e}\mathbf{e}_R}\,, \tag{20}$$

$$\mathrm{SNR}_E^{\mathrm{in}} = \frac{\mathcal{E}\left\{|x_E|^2\right\}}{\mathcal{E}\left\{|n_E|^2\right\}} = \frac{\mathbf{e}_E^T\mathbf{R}_{x,e}\mathbf{e}_E}{\mathbf{e}_E^T\mathbf{R}_{n,e}\mathbf{e}_E}\,. \tag{21}$$

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

$$\mathrm{SNR}_L^{\mathrm{out}} = \frac{\mathbf{w}_L^H\mathbf{R}_{x,e}\mathbf{w}_L}{\mathbf{w}_L^H\mathbf{R}_{n,e}\mathbf{w}_L}\,,\quad \mathrm{SNR}_R^{\mathrm{out}} = \frac{\mathbf{w}_R^H\mathbf{R}_{x,e}\mathbf{w}_R}{\mathbf{w}_R^H\mathbf{R}_{n,e}\mathbf{w}_R}\,. \tag{22}$$

For the binaural hearing device configuration they are defined similarly by replacing $\mathbf{R}_{x,e}$ with $\mathbf{R}_x$ and $\mathbf{R}_{n,e}$ with $\mathbf{R}_n$. The input IC of the desired source component and the noise component is defined as the normalized cross-correlation between the reference microphone signals, i.e.,

$$\mathrm{IC}_x^{\mathrm{in}} = \frac{\mathcal{E}\left\{x_L x_R^*\right\}}{\sqrt{\mathcal{E}\left\{|x_L|^2\right\}\mathcal{E}\left\{|x_R|^2\right\}}} = \frac{\mathbf{e}_L^T\mathbf{R}_{x,e}\mathbf{e}_R}{\sqrt{\left(\mathbf{e}_L^T\mathbf{R}_{x,e}\mathbf{e}_L\right)\left(\mathbf{e}_R^T\mathbf{R}_{x,e}\mathbf{e}_R\right)}}\,, \tag{23}$$

$$\mathrm{IC}_n^{\mathrm{in}} = \frac{\mathcal{E}\left\{n_L n_R^*\right\}}{\sqrt{\mathcal{E}\left\{|n_L|^2\right\}\mathcal{E}\left\{|n_R|^2\right\}}} = \frac{\mathbf{e}_L^T\mathbf{R}_{n,e}\mathbf{e}_R}{\sqrt{\left(\mathbf{e}_L^T\mathbf{R}_{n,e}\mathbf{e}_L\right)\left(\mathbf{e}_R^T\mathbf{R}_{n,e}\mathbf{e}_R\right)}}\,. \tag{24}$$

The output IC of the desired source component and the noise component is defined as the normalized cross-correlation between the output signals, i.e.,

$$\mathrm{IC}_x^{\mathrm{out}} = \frac{\mathbf{w}_L^H\mathbf{R}_{x,e}\mathbf{w}_R}{\sqrt{\left(\mathbf{w}_L^H\mathbf{R}_{x,e}\mathbf{w}_L\right)\left(\mathbf{w}_R^H\mathbf{R}_{x,e}\mathbf{w}_R\right)}}\,, \tag{25}$$

$$\mathrm{IC}_n^{\mathrm{out}} = \frac{\mathbf{w}_L^H\mathbf{R}_{n,e}\mathbf{w}_R}{\sqrt{\left(\mathbf{w}_L^H\mathbf{R}_{n,e}\mathbf{w}_L\right)\left(\mathbf{w}_R^H\mathbf{R}_{n,e}\mathbf{w}_R\right)}}\,. \tag{26}$$

For the binaural hearing device configuration they are defined similarly by replacing $\mathbf{R}_{x,e}$ with $\mathbf{R}_x$ and $\mathbf{R}_{n,e}$ with $\mathbf{R}_n$. The MSC is defined as the square of the absolute value of the IC, i.e.,

$$\mathrm{MSC} = |\mathrm{IC}|^2\,. \tag{27}$$

For the (coherent) desired source, it can be shown using (15) that the input IC in (23) is equal to [21]

$$\mathrm{IC}_x^{\mathrm{in}} = \frac{a_L a_R^*}{|a_L||a_R|} = e^{j\,a_L/a_R}\,, \tag{28}$$

such that the input MSC is equal to 1, i.e.,

$$\mathrm{MSC}_x^{\mathrm{in}} = 1\,. \tag{29}$$

III. BMVDR Beamformers

In this section, we briefly review two binaural noise reduction algorithms using only the head-mounted microphone signals: the BMVDR beamformer and the BMVDR beamformer with partial noise estimation.

A. Binaural Minimum Variance Distortionless Response (BMVDR) Beamformer

The BMVDR beamformer minimizes the PSD of the noise component in the output signals while preserving the desired source component in the left and the right reference microphone signals. The optimization problem for the left and the right filter vector is given by [2], [39], [40]

$$\min_{\mathbf{w}_L}\mathcal{E}\left\{|\mathbf{w}_L^H\mathbf{n}|^2\right\}\quad\text{subject to}\quad \mathbf{w}_L^H\mathbf{x} = x_L\,, \tag{30}$$

$$\min_{\mathbf{w}_R}\mathcal{E}\left\{|\mathbf{w}_R^H\mathbf{n}|^2\right\}\quad\text{subject to}\quad \mathbf{w}_R^H\mathbf{x} = x_R\,. \tag{31}$$

Using (4), (8) and (9), the left and the right filter vector for the BMVDR beamformer are given by [2]

$$\mathbf{w}_{\mathrm{BMVDR},L} = \frac{\mathbf{R}_n^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{R}_n^{-1}\mathbf{a}}a_L^* = \frac{\mathbf{R}_n^{-1}\mathbf{a}_L}{\mathbf{a}_L^H\mathbf{R}_n^{-1}\mathbf{a}_L}\,, \tag{32}$$

$$\mathbf{w}_{\mathrm{BMVDR},R} = \frac{\mathbf{R}_n^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{R}_n^{-1}\mathbf{a}}a_R^* = \frac{\mathbf{R}_n^{-1}\mathbf{a}_R}{\mathbf{a}_R^H\mathbf{R}_n^{-1}\mathbf{a}_R}\,. \tag{33}$$

By substituting (32) and (33) in (22), it has been shown in [3], [9] that the output SNR of the BMVDR beamformer for both the left and the right hearing device is equal to

$$\boxed{\rho = \mathrm{SNR}_{\mathrm{BMVDR},L}^{\mathrm{out}} = \mathrm{SNR}_{\mathrm{BMVDR},R}^{\mathrm{out}} = p_s\mathbf{a}^H\mathbf{R}_n^{-1}\mathbf{a}}\,, \tag{34}$$

which is always larger than or equal to the input SNR in (19) and (20). As can be observed from (34), ρ depends on the ATF vector $\mathbf{a}$, i.e., the position of the head-mounted microphones and the desired source, the noise covariance matrix $\mathbf{R}_n$ and the PSD p_s of the desired source signal.

As has been shown in [3], [9], [21], the BMVDR beamformer preserves the binaural cues of the desired source component, but distorts the binaural cues of the noise component in such a way that both output components exhibit the binaural cues of the desired source component. Hence, the output IC of the noise component for the BMVDR beamformer is equal to [21]

$$\mathrm{IC}^{\mathrm{out}}_{\mathrm{BMVDR},n} = \mathrm{IC}^{\mathrm{in}}_{x} = e^{j\,a_L/a_R}\,, \tag{35}$$

such that the output MSC of the noise component is equal to

$$\mathrm{MSC}^{\mathrm{out}}_{\mathrm{BMVDR},n} = \mathrm{MSC}^{\mathrm{in}}_{x} = 1\,. \tag{36}$$

This means for any arbitrary noise field (e.g., a diffuse noise field) that at the output of the BMVDR beamformer it would be perceived as a coherent source from the direction of the desired source, which is obviously undesired in terms of sound quality and spatial awareness.

B. BMVDR Beamformer With Partial Noise Estimation (BMVDR-N)

Aiming at additionally preserving the binaural cues of the noise component and hence the spatial impression of the complete acoustic scene, the BMVDR-N beamformer has been proposed [3], [9], [16], [18]. The BMVDR-N beamformer 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 left and the right reference microphone signals. The optimization problem for the left and the right filter vector is given by

$$\min_{\mathbf{w}_L} \mathcal{E}\left\{|\mathbf{w}_L^H\mathbf{n} - \eta n_L|^2\right\} \quad \text{subject to} \quad \mathbf{w}_L^H\mathbf{x} = x_L\,, \tag{37}$$

$$\min_{\mathbf{w}_R} \mathcal{E}\left\{|\mathbf{w}_R^H\mathbf{n} - \eta n_R|^2\right\} \quad \text{subject to} \quad \mathbf{w}_R^H\mathbf{x} = x_R\,, \tag{38}$$

where η denotes a (real-valued) mixing parameter, with $0 \leq \eta \leq 1$. It has been shown in [9], [18] that the resulting left and right filter vectors can be written as

$$\mathbf{w}_{\mathrm{BMVDR-N},L} = (1-\eta)\mathbf{w}_{\mathrm{BMVDR},L} + \eta\mathbf{e}_L\,, \tag{39}$$

$$\mathbf{w}_{\mathrm{BMVDR-N},R} = (1-\eta)\mathbf{w}_{\mathrm{BMVDR},R} + \eta\mathbf{e}_R\,, \tag{40}$$

i.e., the output signals of the BMVDR-N beamformer can be interpreted as a mixture of the output signals of the BMVDR beamformer (scaled with $1-\eta$) and the noisy reference microphone signals (scaled with η). For $\eta = 0$, the BMVDR-N beamformer in (39) and (40) is equal to the BMVDR beamformer in (32) and (33). For $\eta = 1$, only the reference microphone signals are used, i.e., no beamforming is applied at all.

In [9], [18] it has been shown that the left and the right output SNR of the BMVDR-N beamformer are equal to

$$\mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR-N},L} = \frac{\rho}{1+\eta^2(\frac{\rho}{\mathrm{SNR}^{\mathrm{in}}_L}-1)}\,, \tag{41}$$

$$\mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR-N},R} = \frac{\rho}{1+\eta^2(\frac{\rho}{\mathrm{SNR}^{\mathrm{in}}_R}-1)}\,. \tag{42}$$

Since $\rho \geq \mathrm{SNR}^{\mathrm{in}}_L$ and $\rho \geq \mathrm{SNR}^{\mathrm{in}}_R$, it can easily be seen that (41) and (42) monotonically decrease with increasing η, such that a larger mixing parameter η leads to a smaller output SNR of the BMVDR-N beamformer [9], [18], i.e.,

$$\mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR-N},L} \leq \mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR},L}\,, \tag{43}$$

$$\mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR-N},R} \leq \mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR},R}\,. \tag{44}$$

Similar to the BMVDR beamformer, the BMVDR-N beamformer preserves the binaural cues of the desired source component [9], [18]. By substituting (39) and (40) in (26) and (27), it has been shown in [18] that the output MSC of the noise component for the BMVDR-N beamformer is equal to

$$\mathrm{MSC}^{\mathrm{out}}_{\mathrm{BMVDR-N},n} = \frac{\left|\frac{1-\eta^2}{\rho}p_s a_L a_R^* + \eta^2 p_{n_{LR}}\right|^2}{\left(\frac{1-\eta^2}{\rho}p_s|a_L|^2 + \eta^2 p_{n_L}\right)\left(\frac{1-\eta^2}{\rho}p_s|a_R|^2 + \eta^2 p_{n_R}\right)}\,, \tag{45}$$

with $p_{n_{LR}} = \mathcal{E}\{n_L n_R^*\}$ the cross power spectral density (CPSD) of the noise component in the reference microphone signals, and $p_{n_L} = \mathcal{E}\{|n_L|^2\}$ and $p_{n_R} = \mathcal{E}\{|n_R|^2\}$ the PSD of the noise component in the left and the right reference microphone signal, respectively.

For $\eta = 1$, the output MSC of the noise component in (45) is equal to the input MSC of the noise component. For $\eta = 0$, the output MSC of the noise component is equal to 1. Since a larger mixing parameter leads to a better MSC preservation of the noise component (i.e., better preservation of the spatial impression) but a smaller output SNR, this mixing parameter allows to trade off between noise reduction performance and binaural cue preservation of the noise component.

In order to achieve a desired output MSC $\mathrm{MSC}^{\mathrm{des}}_n$ of the noise component, with

$$0 \leq \mathrm{MSC}^{\mathrm{in}}_n \leq \mathrm{MSC}^{\mathrm{des}}_n \leq 1\,, \tag{46}$$

in [18] a closed-form expression for the mixing parameter η^{des} has been derived, i.e.,

$$\boxed{\eta^{\mathrm{des}} = \sqrt{\frac{\rho\left(\sqrt{\gamma^2-\alpha\beta}-\gamma\right)+\alpha}{\rho^2\beta-2\rho\gamma+\alpha}}}\,, \tag{47}$$

with

$$\alpha = \left(\mathrm{MSC}^{\mathrm{des}}_n - 1\right)p_s^2|a_L|^2|a_R|^2\,, \tag{48}$$

$$\beta = \left(\mathrm{MSC}^{\mathrm{des}}_n - \mathrm{MSC}^{\mathrm{in}}_n\right)p_{n_L}p_{n_R}\,, \tag{49}$$

$$\gamma = \mathrm{MSC}^{\mathrm{des}}_n\frac{p_s|a_L|^2 p_{n_L} + p_s|a_R|^2 p_{n_R}}{2} - \Re\{p_s a_L a_R^* p_{n_{LR}}^*\}\,, \tag{50}$$

with $\Re\{\cdot\}$ the real part of a complex number. Since $\mathrm{MSC}^{\mathrm{in}}_n \leq \mathrm{MSC}^{\mathrm{des}}_n \leq 1$ and all PSDs are positive (or zero), it can be easily seen that $\alpha \leq 0$ and $\beta \geq 0$. Aiming for the spatial impression of the noise component in the reference microphone signals and the noise component in the output signals to be indistinguishable, it has been proposed in [18], [21] to define the desired output MSC of the noise component based on the IC discrimination ability of the human auditory system [41], [42].

IV. Extended BMVDR Beamformers Incorporating an External Microphone

In this section, we consider the BMVDR and BMVDR-N beamformers for the *extended* binaural hearing device configuration, i.e., using an additional external microphone in conjunction with the head-mounted microphones. In Section IV-A, we define the *extended* BMVDR (eBMVDR) beamformer and the *extended* BMVDR-N (eBMVDR-N) beamformer. In the following sections we derive analytical expressions for the output SNR and the output MSC of the noise component for the eBMVDR-N beamformer for an arbitrary noise field and without assuming a specific position of the desired source. In Section IV-B we show that 1) the same mixing parameter η leads to a larger (or equal) output SNR for the eBMVDR-N beamformer than for the BMVDR-N beamformer, and 2) the same output SNR can be obtained for a larger (or equal) mixing parameter in the eBMVDR-N beamformer than in the BMVDR-N beamformer. In Section IV-C we show that the same desired output MSC of the noise component can be obtained for a smaller (or equal) mixing parameter in the eBMVDR-N beamformer than in the BMVDR-N beamformer. These results generalize the results obtained in [23] assuming a coherent (directional) noise source, and the results in [31] assuming a homogeneous noise field and a desired source in front of the listener.

A. Extended BMVDR and BMVDR-N Beamformers

The BMVDR beamformer incorporating the external microphone is referred to as the *extended* BMVDR (eBMVDR) beamformer. Similarly to (32) and (33), by replacing the noise covariance matrix $\mathbf{R}_n$ with the extended noise covariance matrix $\mathbf{R}_{n,e}$ in (16) and the RTF vectors $\mathbf{a}_L$ and $\mathbf{a}_R$ with the extended RTF vectors $\mathbf{a}_{L,e}$ and $\mathbf{a}_{R,e}$ in (14) the left and the right filter vector of the eBMVDR beamformer are equal to

$$\mathbf{w}_{\text{eBMVDR},L} = \frac{\mathbf{R}_{n,e}^{-1}\mathbf{a}_e}{\mathbf{a}_e^H\mathbf{R}_{n,e}^{-1}\mathbf{a}_e}a_L^* = \frac{\mathbf{R}_{n,e}^{-1}\mathbf{a}_{L,e}}{\mathbf{a}_{L,e}^H\mathbf{R}_{n,e}^{-1}\mathbf{a}_{L,e}}\,, \tag{51}$$

$$\mathbf{w}_{\text{eBMVDR},R} = \frac{\mathbf{R}_{n,e}^{-1}\mathbf{a}_e}{\mathbf{a}_e^H\mathbf{R}_{n,e}^{-1}\mathbf{a}_e}a_R^* = \frac{\mathbf{R}_{n,e}^{-1}\mathbf{a}_{R,e}}{\mathbf{a}_{R,e}^H\mathbf{R}_{n,e}^{-1}\mathbf{a}_{R,e}}\,. \tag{52}$$

The BMVDR beamformer with partial noise estimation incorporating the external microphone signal is referred to as the *extended* BMVDR-N (eBMVDR-N) beamformer. Similarly to (39) and (40), the left and the right filter vector of the eBMVDR-N beamformer are equal to

$$\mathbf{w}_{\text{eBMVDR-N},L} = (1-\eta)\mathbf{w}_{\text{eBMVDR},L} + \eta\mathbf{e}_L\,, \tag{53}$$

$$\mathbf{w}_{\text{eBMVDR-N},R} = (1-\eta)\mathbf{w}_{\text{eBMVDR},R} + \eta\mathbf{e}_R\,, \tag{54}$$

where η again denotes the (real-valued) mixing parameter, with $0 \leq \eta \leq 1$. The output signals of the eBMVDR-N beamformer are again equal to a mixture of the output signals of the eBMVDR beamformer (scaled with $1-\eta$) and the (noisy) reference microphone signals (scaled with η).

Similarly to the mixing parameter η^{des} for the BMVDR-N beamformer in (47), the mixing parameter η_e^{des} for the eBMVDR-N beamformer leading to a desired output MSC $\text{MSC}_n^{\text{des}}$ of the noise component is equal to

$$\boxed{\eta_e^{\text{des}} = \sqrt{\frac{\rho_e\left(\sqrt{\gamma^2-\alpha\beta}-\gamma\right)+\alpha}{\rho_e^2\beta-2\rho_e\gamma+\alpha}}}, \tag{55}$$

with α, β and γ defined in (48)–(50) and ρ_e defined in (56).

B. Output SNR With an External Microphone

Similarly to (34) by substituting (53) and (54) in (22), the output SNR of the eBMVDR beamformer is equal to [31]

$$\boxed{\rho_e = \text{SNR}_{\text{eBMVDR},L}^{\text{out}} = \text{SNR}_{\text{eBMVDR},R}^{\text{out}} = p_s\mathbf{a}_e^H\mathbf{R}_{n,e}^{-1}\mathbf{a}_e}. \tag{56}$$

The inverse of the extended noise covariance matrix $\mathbf{R}_{n,e}$ can be written in terms of $\mathbf{R}_n^{-1}$ as [43]

$$\mathbf{R}_{n,e}^{-1} = \left[\begin{array}{c|c}\mathbf{R}_n^{-1}+\frac{1}{\xi}\mathbf{R}_n^{-1}\mathbf{r}_{n,E}\mathbf{r}_{n,E}^H\mathbf{R}_n^{-1} & -\frac{1}{\xi}\mathbf{R}_n^{-1}\mathbf{r}_{n,E} \\ \hline -\frac{1}{\xi}\mathbf{r}_{n,E}^H\mathbf{R}_n^{-1} & \frac{1}{\xi}\end{array}\right], \tag{57}$$

with $\xi = p_{n_E} - \mathbf{r}_{n,E}^H\mathbf{R}_n^{-1}\mathbf{r}_{n,E}$ the Schur complement of $\mathbf{R}_n$ in (16). It can be shown that $\xi > 0$, since $\mathbf{R}_{n,e}$ is assumed to be positive definite.

By substituting (57) in (56) and using (34), the output SNR of the eBMVDR beamformer can be written as

$$\begin{aligned}\rho_e &= p_s\left(\mathbf{a}^H\mathbf{R}_n^{-1}\mathbf{a} + \frac{1}{\xi}\left|\mathbf{r}_{n,E}^H\mathbf{R}_n^{-1}\mathbf{a}-a_E\right|^2\right) \\ &= \rho + p_s\frac{\left|\mathbf{r}_{n,E}^H\mathbf{R}_n^{-1}\mathbf{a}-a_E\right|^2}{p_{n_E}-\mathbf{r}_{n,E}^H\mathbf{R}_n^{-1}\mathbf{r}_{n,E}}\,.\end{aligned} \tag{58}$$

Hence, as expected, the output SNR ρ_e of the eBMVDR beamformer is always larger than or equal to the output SNR ρ of the BMVDR beamformer (without an external microphone), i.e.,

$$\boxed{\rho_e \geq \rho}. \tag{59}$$

As can be observed from (58), the SNR improvement due to incorporating the external microphone depends on the ATF a_E between the desired source and the external microphone. In addition, the SNR improvement depends on the PSD p_{n_E} of the noise component in the external microphone signal and the spatial correlation $\mathbf{r}_{n,E}$ between the noise component in the head-mounted microphones signals and the external microphone signal. This implies that the SNR improvement obviously depends on the position of the external microphone relative to the head-mounted microphones and the desired source.

Similarly to (41) and (42), the left and the right output SNR of the eBMVDR-N beamformer are equal to

$$\text{SNR}_{\text{eBMVDR-N},L}^{\text{out}} = \frac{\rho_e}{1+\eta^2(\frac{\rho_e}{\text{SNR}_L^{\text{in}}}-1)}\,, \tag{60}$$

$$\text{SNR}_{\text{eBMVDR-N},R}^{\text{out}} = \frac{\rho_e}{1+\eta^2(\frac{\rho_e}{\text{SNR}_R^{\text{in}}}-1)}\,. \tag{61}$$

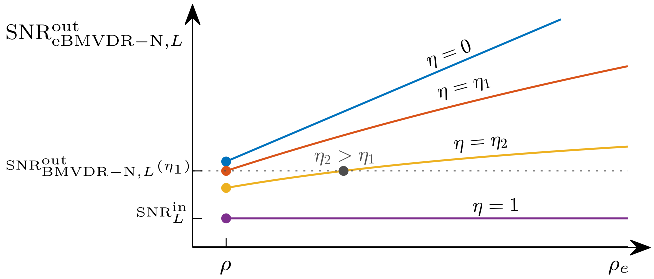


Fig. 3. Left output SNR of the eBMVDR-N beamformer as a function of the output SNR ρ_e of the eBMVDR beamformer for different values of the mixing parameter η. Please note that $\eta = 0$ corresponds to the eBMVDR beamformer and $\eta = 1$ corresponds to the left reference microphone signal.

Since $\rho_e \geq \mathrm{SNR}^{\mathrm{in}}_L$ and $\rho_e \geq \mathrm{SNR}^{\mathrm{in}}_R$, (60) and (61) also monotonically decrease with increasing η, such that a larger mixing parameter η leads to a smaller output SNR of the eBMVDR-N beamformer. Fig. 3 depicts the left output SNR of the eBMVDR-N beamformer in (60) as a function of the output SNR ρ_e of the eBMVDR beamformer for different values of the mixing parameter η. For a given binaural hearing device configuration (i.e., positions of the head-mounted microphones), desired source position and noise field, the output SNR ρ in (34) of the BMVDR beamformer is a constant.

Now consider different positions of the external microphone, such that the output SNR ρ_e of the eBMVDR beamformer in (58) can be considered as a variable. Based on (59), the smallest possible value for the output SNR ρ_e of the eBMVDR beamformer is equal to ρ, i.e., the output SNR of the BMVDR beamformer (without an external microphone). For a fixed value of the mixing parameter η, it can be easily shown that the partial derivative of (60) with respect to ρ_e is equal to

$$\frac{\partial \mathrm{SNR}^{\mathrm{out}}_{\mathrm{eBMVDR-N},L}}{\partial \rho_e} = \frac{1-\eta^2}{\left(1+\eta^2\left(\frac{\rho_e}{\mathrm{SNR}^{\mathrm{in}}_L}-1\right)\right)^2} \geq 0\,, \quad (62)$$

since $0 \leq \eta \leq 1$. Hence, for each value of the mixing parameter, e.g., $\eta = \eta_1$ (see Fig. 3), the left output SNR of the eBMVDR-N beamformer monotonically increases with ρ_e, and using (59), is always larger than or equal to the left output SNR of the BMVDR-N beamformer, i.e.,

$$\boxed{\mathrm{SNR}^{\mathrm{out}}_{\mathrm{eBMVDR-N},L}(\eta_1) \geq \mathrm{SNR}^{\mathrm{out}}_{\mathrm{BMVDR-N},L}(\eta_1)}\,. \quad (63)$$

In addition, for any $\rho_e > \rho$ the same output SNR of the BMVDR-N beamformer can only be obtained when using a larger mixing parameter $\eta_2 > \eta_1$ for the eBMVDR-N beamformer.

This means that incorporating an external microphone allows to use a larger mixing parameter, i.e., achieve a better spatial impression of the noise component, to obtain the same output SNR compared to only using the head-mounted microphone signals. Similar expressions can be derived for the right output SNR of the eBMVDR-N beamformer in (61).

C. Output MSC With an External Microphone

As discussed in Section III-B, the mixing parameter η controls the binaural cues of the noise component at the output of the BMVDR-N beamformer. Since a larger mixing parameter leads to a lower output SNR, it is hence desirable to achieve the desired binaural cues of the noise output component using a small mixing parameter. For the special case of a coherent (directional) noise source, it has been experimentally shown in [23] that the same binaural cues, i.e., ILD and ITD, of the output noise component can be achieved using a smaller mixing parameter when incorporating an external microphone compared to using only the head-mounted microphones. Further, for the special case of a homogeneous noise field and a desired source in front of the listener, it has been analytically shown in [31] that the same desired output MSC of the noise component can be achieved using a smaller mixing parameter in the eBMVDR-N beamformer than in the BMVDR-N beamformer. In this section, we generalize the analytical expressions derived in [31] without making any assumption about the noise field and the position of the desired source.

Since it was not straightforward to directly prove that η^{des}_e in (55) is always smaller than (or equal to) η^{des} in (47), we will take an indirect approach. Since $\rho_e \geq \rho$, showing that $\eta^{\mathrm{des}}_e \leq \eta^{\mathrm{des}}$ corresponds to showing that η^{des}_e monotonically decreases with ρ_e.

The squared mixing parameter in (55) can be written as

$$\left(\eta^{\mathrm{des}}_e\right)^2 = \frac{\nu_1(\rho_e)}{\nu_2(\rho_e)}\,, \quad (64)$$

with

$$\nu_1(\rho_e) = \rho_e(\kappa-\gamma)+\alpha\,, \quad (65)$$

$$\nu_2(\rho_e) = \rho_e^2\beta - 2\rho_e\gamma + \alpha\,, \quad (66)$$

with

$$\kappa = \sqrt{\gamma^2-\alpha\beta}\,, \quad (67)$$

and with α, β and γ defined in (48)–(50). Using the quotient rule to compute the partial derivative of (64) with respect to ρ_e gives

$$\frac{\partial(\eta^{\mathrm{des}}_e)^2}{\partial\rho_e} = 2\eta^{\mathrm{des}}_e\frac{\partial\eta^{\mathrm{des}}_e}{\partial\rho_e} = \frac{\frac{\partial\nu_1(\rho_e)}{\partial\rho_e}\nu_2(\rho_e) - \frac{\partial\nu_2(\rho_e)}{\partial\rho_e}\nu_1(\rho_e)}{\nu_2^2(\rho_e)}\,. \quad (68)$$

Hence, since $\eta^{\mathrm{des}}_e \geq 0$ and $\nu_2^2(\rho_e) \geq 0$, in order to show that η^{des}_e monotonically decreases with ρ_e, i.e.,

$$\frac{\partial\eta^{\mathrm{des}}_e}{\partial\rho_e} \leq 0\,, \quad (69)$$

it is sufficient to show that

$$\zeta(\rho_e) = \frac{\partial\nu_1(\rho_e)}{\partial\rho_e}\nu_2(\rho_e) - \frac{\partial\nu_2(\rho_e)}{\partial\rho_e}\nu_1(\rho_e) \leq 0\,. \quad (70)$$

Computing the partial derivatives of (65) and (66) with respect to ρ_e gives

$$\frac{\partial\nu_1(\rho_e)}{\partial\rho_e} = \kappa-\gamma\,, \quad (71)$$

$$\frac{\partial\nu_2(\rho_e)}{\partial\rho_e} = 2\rho_e\beta - 2\gamma\,. \quad (72)$$

By substituting (65), (66), (71) and (72) in (70), $\zeta(\rho_e)$ can be written as a quadratic function of ρ_e, i.e.,

$$\zeta(\rho_e) = \psi_1 \rho_e^2 - 2\psi_2 \rho_e + \psi_3 \,, \tag{73}$$

with

$$\psi_1 = (\gamma - \kappa)\beta \,, \tag{74}$$

$$\psi_2 = \alpha\beta = \gamma^2 - \kappa^2 \,, \tag{75}$$

$$\psi_3 = \alpha(\kappa + \gamma) \,. \tag{76}$$

The extremum of the quadratic function $\zeta(\rho_e)$ in (73) can be found by setting $\frac{\partial \zeta(\rho_e)}{\partial \rho_e} = 0$, leading to

$$\tilde{\rho}_e = \frac{\psi_2}{\psi_1} = \frac{\alpha}{\gamma - \kappa} = \frac{\kappa + \gamma}{\beta} \,. \tag{77}$$

Substituting (77) in (73) yields

$$\zeta(\tilde{\rho}_e) = \frac{\psi_1 \psi_3 - \psi_2^2}{\psi_1} = 0 \,. \tag{78}$$

The second-order partial derivative of $\zeta(\rho_e)$ in (73) is equal to

$$\frac{\partial^2 \zeta(\rho_e)}{\partial \rho_e^2} = 2\psi_1 = 2(\gamma - \kappa)\beta \,. \tag{79}$$

Since $\alpha \leq 0$ and $\beta \geq 0$ (see Section III-B), using (67) it follows that

$$\alpha\beta = (\gamma - \kappa)(\gamma + \kappa) \leq 0 \,. \tag{80}$$

We now consider two cases:

1) $\gamma \geq 0$: Since $\kappa \geq 0$, it follows that $\gamma + \kappa \geq 0$, such that $\gamma - \kappa \leq 0$ in order to satisfy (80).
2) $\gamma \leq 0$: Since $\kappa \geq 0$, it directly follows that $\gamma - \kappa \leq 0$.

Since $\gamma - \kappa \leq 0$ and $\beta \geq 0$, the second-order partial derivative in (79) is always negative (or equal to zero). Since the extremum is hence a maximum with function value 0, cf. (78), the quadratic function $\zeta(\rho_e)$ in (73) is negative (or zero) for all values of ρ_e. Hence, $\frac{\partial \eta_e^{\text{des}}}{\partial \rho_e} \leq 0$, such that

$$\boxed{\eta_e^{\text{des}} \leq \eta^{\text{des}}} \,, \tag{81}$$

i.e., to achieve the same desired output MSC of the noise component a smaller mixing parameter can be used in the eBMVDR-N beamformer (incorporating an external microphone) than in the BMVDR-N beamformer (using only the head-mounted microphones). Together with the SNR results obtained in Section IV-B, this implies that for any arbitrary noise field and position of the desired source an external microphone enables the same spatial impression of the noise component to be achieved while achieving a larger output SNR.

V. Experimental Results

In Section V-A we first validate the analytical expressions derived in the previous sections using simulated anechoic ATFs for various positions of the external microphone. In Section V-B we provide experimental results using recorded signals in a reverberant environment with multiple interfering speakers as background noise, showing that also in a realistic scenario incorporating an external microphone enables the output SNR to be significantly increased and the mixing parameter required to obtain a desired output MSC of the noise component to be decreased.

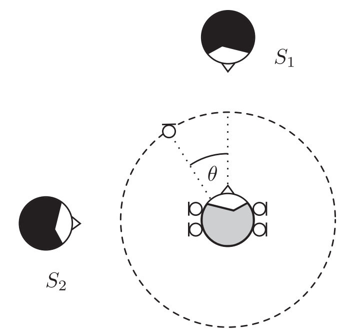


Fig. 4. Anechoic validation setup using 2 microphones on each side of the head. The external microphone was placed at 3 m distance to the listener for different angles θ. The desired source was placed at 3.5 m distance at two different angles, i.e., S_1 at 0° and S_2 at −90°.

A. Validation Using Anechoic ATFs

To validate our theoretical findings from Section IV, we simulated an anechoic acoustic scenario, where the head of the listener was modelled as a rigid sphere with a diameter of 17 cm [32]. Without considering any hearing devices, we considered 2 microphones on each side of the head, i.e., $M_H = 4$, with an inter-microphone distance of 7 mm, such that including the external microphone the total number of microphones was $M = 5$.

Fig. 4 depicts the validation setup. The external microphone was placed at a distance of 3 m to the listener, where the azimuth angle θ was varied from $-180°$ to $180°$. The desired source was placed at a distance of 3.5 m to the listener at two different angles, i.e., S_1 at 0° (in front) and S_2 at $-90°$ (to the left). Hence, the smallest distance between the external microphone and the desired source was equal to 0.5 m, whereas the largest distance was equal to 6.5 m.

All ATFs were simulated at a sampling rate of 16 kHz using an FFT length of 256 samples. For the ATFs $\mathbf{a}$ (corresponding to the head-mounted microphones) we used the SMIR generator [32] to consider a rigid sphere, while for the ATF $\mathbf{a}_e$ (corresponding to the external microphone) we set a parameter of the generator to ignore the sphere. For the speech PSD we assumed a flat spectrum, i.e. $p_s = 1$. As background noise, we considered 8 mutually independent noise sources with equal power at angles $\{-140°, -100°, -60°, -20°, 20°, 60°, 100°, 140°\}$, resulting in a diffuse-like noise field which is neither coherent nor perfectly diffuse. The extended noise covariance matrix $\mathbf{R}_{n,e}$ was calculated as the sum of the 8 corresponding (rank-1) covariance matrices, constructed using the simulated ATF vectors of the noise sources. As reference microphones we considered the front microphones on the left side and the right side. The input SNR in the left reference microphone signal was set to

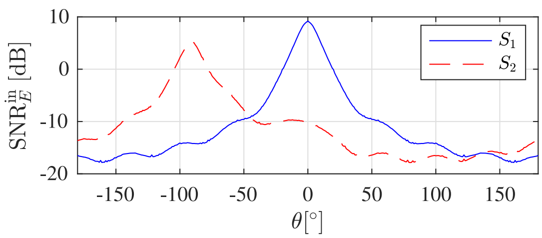


Fig. 5. Input SNR in the external microphone signal (averaged over all frequencies) for different angles θ of the external microphone for both considered positions of the desired source S_1 and S_2.

0 dB (averaged over all frequencies), leading to the input SNR in the external microphone signal (averaged over all frequencies) as depicted in Fig. 5. As can be observed, the input SNR in the external microphone signal varied within a range of nearly 30 dB, with the highest input SNR occuring when the external microphone is closest to the desired source (i.e., 0° for S_1 and $-90°$ for S_2).

a) BMVDR beamformer vs. eBMVDR beamformer: In Section IV, we showed that the output SNR of the eBMVDR beamformer ρ_e in (56) is always larger than or equal to the output SNR of the BMVDR beamformer ρ in (34). Fig. 6 depicts the benefit of incorporating an external microphone in terms of the output SNR ratio, i.e.,

$$\frac{\rho_e}{\rho} = \frac{\mathbf{a}_e^H \mathbf{R}_{n,e}^{-1} \mathbf{a}_e}{\mathbf{a}^H \mathbf{R}_n^{-1} \mathbf{a}} \tag{82}$$

for different angles of the external microphone and for both considered positions of the desired source. As can be observed, for all positions of the external microphone and the desired source and for all frequencies $\rho_e \geq \rho$, hence satisfying (59). Moreover, the benefit of incorporating the external microphone is larger for small distances between the desired source and the external microphone, in this case leading to an SNR improvement of more than 12 dB.

b) BMVDR-N beamformer vs. eBMVDR-N beamformer: In Section IV, we showed that to achieve the same desired output MSC of the noise component the mixing parameter η_e^{des} in (55) of the eBMVDR-N beamformer is smaller than (or equal to) the mixing parameter η^{des} in (47) of the BMVDR-N beamformer, hence leading to a larger output SNR. Here, we set the desired output MSC of the noise component equal to

$$\mathrm{MSC}_n^{\mathrm{des}} = \min\left(1,\, \mathrm{MSC}_n^{\mathrm{in}} + 0.3\right), \tag{83}$$

hence limiting the MSC error to 0.3 for all frequencies and satisfying (46). Fig. 7 depicts the input MSC of the noise component, computed using (24) and (27), and the desired output MSC of the noise component in (83). It can be observed that the input MSC of the noise component resembles a squared (modified) sinc function, as expected for a diffuse-like noise field and modelling the head as a rigid sphere [44]. Fig. 8 depicts the benefit of incorporating the external microphone in terms of the difference between the mixing parameters, i.e., $\eta^{\mathrm{des}} - \eta_e^{\mathrm{des}}$, both leading to the same desired output MSC of the noise component. As can be observed, for all positions of the external microphone and the desired source and for all frequencies $\eta_e^{\mathrm{des}} \leq \eta^{\mathrm{des}}$. As proven in Section IV-C, by comparing Fig. 6 and Fig. 8 it can be observed that a larger output SNR ρ_e of the eBMVDR beamformer leads to a smaller mixing parameter η_e^{des} of the eBMVDR-N beamformer and hence an improved trade-off compared to the mixing parameter η^{des} of the BMVDR-N beamformer. In other words, one has to mix less with the noisy reference microphone signals for the eBMVDR-N beamformer than for the BMVDR-N beamformer, while both beamformers lead to the same spatial impression of the noise component. This effect is larger when the external microphone is close to the desired source, leading to mixing parameter differences that are larger than 0.5.

B. Experimental Results Using Reverberant Recordings

For a more realistic evaluation we used a database with recorded signals in a real-world reverberant environment [33]. The experimental setup is depicted in Fig. 9, where a listener and two different speakers were sitting at a circular table with a diameter of 106 cm in a room with size 12.7 × 10 × 3.6 m^3 and a reverberation time of approximately 620 ms. The setup was surrounded by three layers of in total 56 seated persons producing realistic multi-talker babble noise. The 8 nearest persons were seated about 1 m from the table, with 16 persons at a slightly larger distance of about 2.8 m and 32 persons at a distance of about 4.5 m [33].

Hence, the noise component was diffuse-like, but also contained temporally coherent sources and sensor noise (from the microphones and the recording equipment). The (male) speaker S_1 sat in front of the listener at the other end of the table, while the (female) speaker S_2 sat to the right of the listener. The listener was wearing $M_H = 4$ head-mounted hearing aid microphones, i.e., two microphones on each side. For the external microphone we selected several realistic positions from the database, e.g., representing the listener's smartphone on the table (P_1), a microphone of a conference system in the center of the table (P_2), the smartphone of each speaker placed on the table ($P_{3,1}$ for speaker S_1 and $P_{3,2}$ for speaker S_2) and a headset worn by each speaker ($P_{4,1}$ for speaker S_1 and $P_{4,2}$ for speaker S_2). Only one speaker was active at a time and read 12 sentences for approximately 25 s, while the listener tried to sit as still as possible (but small movements occurred).

We used separate recordings of the speakers and the background noise and mixed them at an intelligibility-weighted SNR (iSNR) for the right reference microphone signal $\mathrm{iSNR}_R^{\mathrm{in}} = 0$ dB. The iSNR is computed by weighting the SNR in each frequency band with a function that takes into account the importance of each frequency band for speech intelligibility [45], [46]. We used a sample rate of 16 kHz and processed the signals in a weighted overlap-add framework using frames of 64 ms with 50% overlap and a square-root Hann window. For the extended BMVDR beamformers using all microphones (i.e., eBMVDR in (51) and (52), and eBMVDR-N in (53) and (54)) the extended noisy input covariance matrix $\mathbf{R}_{y,e}$ and the extended noise covariance matrix $\mathbf{R}_{n,e}$ were recursively

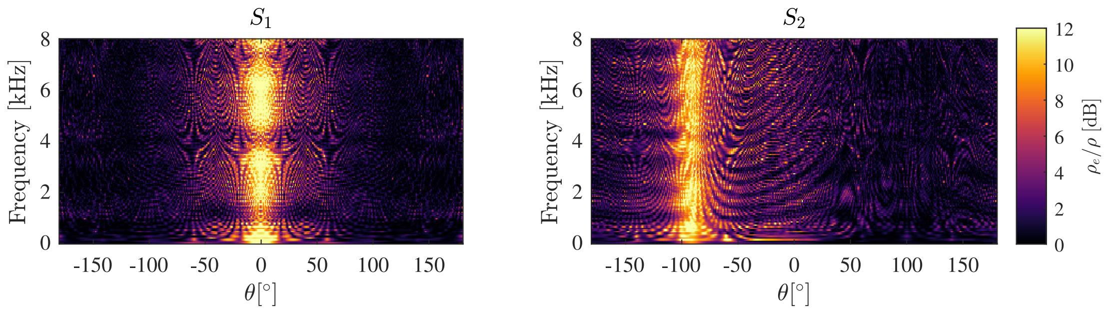


Fig. 6. Benefit of incorporating an external microphone in terms of output SNR (ρ_e/ρ) for different angles θ of the external microphone for (left) position S_1 and (right) position S_2.

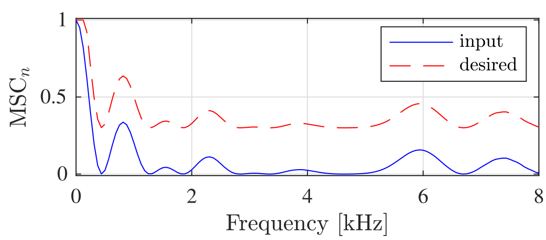


Fig. 7. Input MSC and desired output MSC of the noise component, limiting the MSC error to 0.3.

averaged using a forgetting factor of 0.9, corresponding to approximately 300 ms. To distinguish between speech-plus-noise and noise-only time-frequency bins we used a thresholded speech presence probability (SPP) estimate in the right reference microphone signal, where we used the SPP estimation method proposed in [47]. An SPP estimate larger than or equal to 0.5 was classified as speech-plus-noise, whereas an SPP estimate smaller than 0.5 was classified as noise-only. The (time-varying) extended RTF vectors $\mathbf{a}_{L,e}$ and $\mathbf{a}_{R,e}$ were estimated from the estimated extended covariance matrices $\mathbf{R}_{y,e}$ and $\mathbf{R}_{n,e}$ using the covariance whitening method, i.e., as the principal eigenvector of the pre-whitened extended noisy input covariance matrix $\mathbf{R}_{n,e}^{-1}\mathbf{R}_{y,e}$ [48]–[52]. For the BMVDR beamformers using only the head-mounted microphones (i.e., BMVDR in (32) and (33), and BMVDR-N in (39) and (40)), the covariance matrices $\mathbf{R}_y$ and $\mathbf{R}_n$ were constructed by discarding the last row and the last column of $\mathbf{R}_{y,e}$ and $\mathbf{R}_{n,e}$, respectively. The (time-varying) RTF vectors $\mathbf{a}_L$ and $\mathbf{a}_R$ were estimated from the estimated covariance matrices $\mathbf{R}_y$ and $\mathbf{R}_n$, also using the covariance whitening method.

For the beamformers with partial noise estimation (BMVDR-N and eBMVDR-N), the desired output MSC of the noise component $\mathrm{MSC}_n^{\mathrm{des}}$ was set in a psycho-acoustically motivated way by constraining the output MSC of the noise component by means of frequency-dependent lower and upper boundaries such that the listener's spatial impression of a diffuse noise field should not be altered [17], [18], [21]. These boundaries were defined based on the IC discrimination ability of the human auditory system in diffuse noise fields [41], [42]. Below 500 Hz, the MSC boundaries were chosen as a function of the desired output MSC of the noise component $\mathrm{MSC}_n^{\mathrm{des}}$, whereas above 500 Hz as a fixed lower MSC boundary of 0 and fixed upper MSC boundary of 0.36. Fig. 10 depicts the frequency-dependent long-term input MSC of the noise component and the (psycho-acoustically motivated) frequency-dependent desired output MSC $\mathrm{MSC}_n^{\mathrm{des}}$ of the noise component. Similarly as in Fig. 7, it can be observed that the input MSC of the noise component resembles a squared (modified) sinc function. By re-writing (47) and (55) using the (extended) RTF vectors in (8) and (14), the mixing parameters η^{des} and η_e^{des} of the BMVDR-N and eBMVDR-N beamformers can in practice be computed based on the estimated (extended) RTF vectors and the estimated (extended) noise covariance matrix.

To evaluate the algorithms in terms of noise reduction performance and preservation of the spatial impression of the noise component, based on the output signals we used the left and right iSNR improvement, relating the left and right output iSNR to the left and right input iSNR, i.e.,

$$\Delta\mathrm{iSNR}_L = \mathrm{iSNR}_L^{\mathrm{out}} - \mathrm{iSNR}_L^{\mathrm{in}}, \tag{84}$$

$$\Delta\mathrm{iSNR}_R = \mathrm{iSNR}_R^{\mathrm{out}} - \mathrm{iSNR}_R^{\mathrm{in}}, \tag{85}$$

the modified binaural short-time objective intelligibility (MBSTOI) [53], and the broad-band MSC error $\Delta\mathrm{MSC}_n$ of the noise component, i.e.,

$$\Delta\mathrm{MSC}_n = \frac{1}{F}\sum_{f=1}^{F} \left|\mathrm{MSC}_n^{\mathrm{out}}(f) - \mathrm{MSC}_n^{\mathrm{in}}(f)\right|, \tag{86}$$

where f is the frequency bin index and F is the total number of frequency bins.

Fig. 11 depicts the iSNR improvement, MBSTOI scores, and the MSC error of the noise component for the BMVDR beamformer, the BMVDR-N beamformer, the external microphone signal, the eBMVDR beamformer and the eBMVDR-N beamformer, for the different positions of the external microphone. The top row shows the results for speaker S_1, while the bottom row shows the results for speaker S_2. Considering the iSNR improvements for speaker S_1, it can be observed that incorporating the external microphone signal in the binaural noise reduction algorithms is beneficial for all positions of the external microphone. In terms of iSNR improvement, the

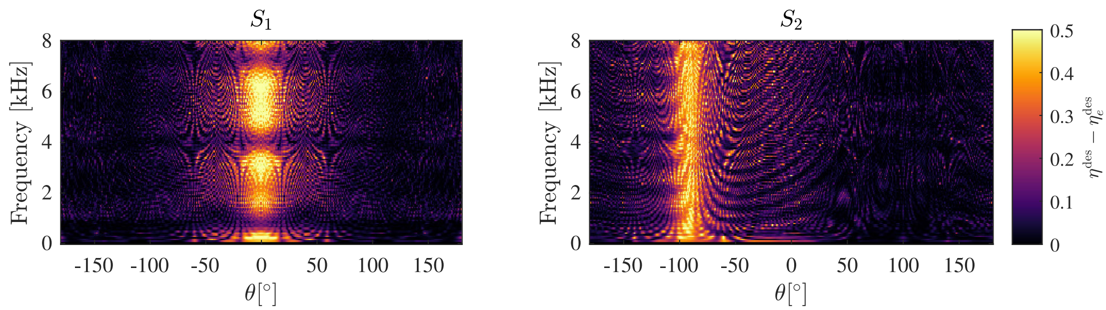


Fig. 8. Difference between the mixing parameter η^{des} of the BMVDR-N beamformer and the mixing parameter η_e^{des} of the eBMVDR-N beamformer, leading to the same desired output MSC of the noise component, for different angles θ of the external microphone for (left) position S_1 and (right) position S_2.

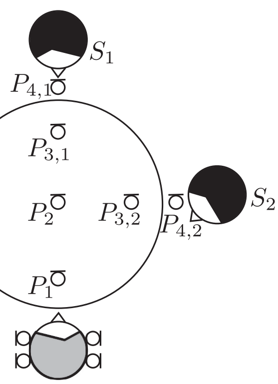


Fig. 9. Experimental realistic setup with a listener wearing head-mounted hearing aid microphones, two different speaker positions (S_1 and S_2) and several possible positions of the external microphone. The setup was surrounded by 56 persons producing realistic multi-talker babble noise.

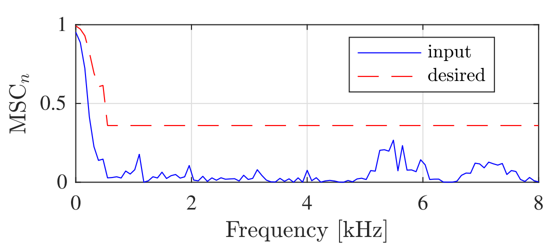


Fig. 10. Measured input MSC of the noise component and (psycho-acoustically motivated) desired output MSC of the noise component.

eBMVDR beamformer always outperforms the BMVDR beamformer and the eBMVDR-N beamformer always outperforms the BMVDR-N beamformer. The iSNR improvement is similar in both hearing devices due to the symmetric scenario for speaker S_1. As expected, the iSNR improvement increases for decreasing distance between the external microphone and speaker S_1 with a very large iSNR improvement for position P_4 (headset microphone). The eBMVDR beamformer outperforms the external microphone signal for all considered positions, whereas the eBMVDR-N beamformer outperforms the external microphone signal for all considered positions except for position P_4. In contrast, the BMVDR beamformer and the BMVDR-N beamformer outperform the external microphone signal only for positions P_1 and P_2, i.e., for the position close to the listener and in the center of the table. Comparing the eBMVDR beamformer to the eBMVDR-N beamformer, it appears that the drop in iSNR improvement for the eBMVDR-N beamformer due to mixing with the noisy reference microphone signals is approximately the same for all positions of the external microphone.

Considering the MBSTOI scores for speaker S_1, it can be observed that the eBMVDR-N beamformer leads to the highest scores for all positions of the external microphone. Partially preserving the MSC of the noise component, i.e., using the BMVDR-N beamformer or the eBMVDR-N beamformer, significantly increases the MBSTOI score compared to not preserving the MSC of the noise component, i.e., using the BMVDR beamformer or the eBMVDR beamformer. This implies that even though the BMVDR-N and eBMVDR-N beamformer lead to a drop in iSNR improvement compared to the BMVDR and BMVDR-N beamformer, the speech intelligibility is increased. These results can be explained by binaural listening effects, e.g., spatial release from masking [54], and are in line with perceptual listening tests in [55]. Since the external microphone signal does not include any binaural cues, the human auditory system cannot benefit from binaural listening effects and the MBSTOI score of the external microphone signal is lower than for all other algorithms (except for position P_4 very close to the speaker).

Considering the MSC error $\Delta\mathrm{MSC}_n$ of the noise component for speaker S_1, as expected only the binaural noise reduction algorithms with partial noise estimation, i.e., the BMVDR-N beamformer and the eBMVDR-N beamformer, are able to yield a low MSC error and hence preserve the spatial impression of the noise component. The external microphone signal obviously shows the worst performance in terms of MSC error of the noise component since the external microphone signal is just a monaural signal that does not include any binaural cues, hence leading to in-head localization.

Considering the results for speaker S_2 (bottom row), it can be observed that similar results as for speaker S_1 are obtained. However, due to the asymmetric setup, the iSNR improvement at the right side (better ear with larger input iSNR) is always smaller than the iSNR improvement at the left side. In addition, the drop in iSNR improvement for the eBMVDR-N beamformer

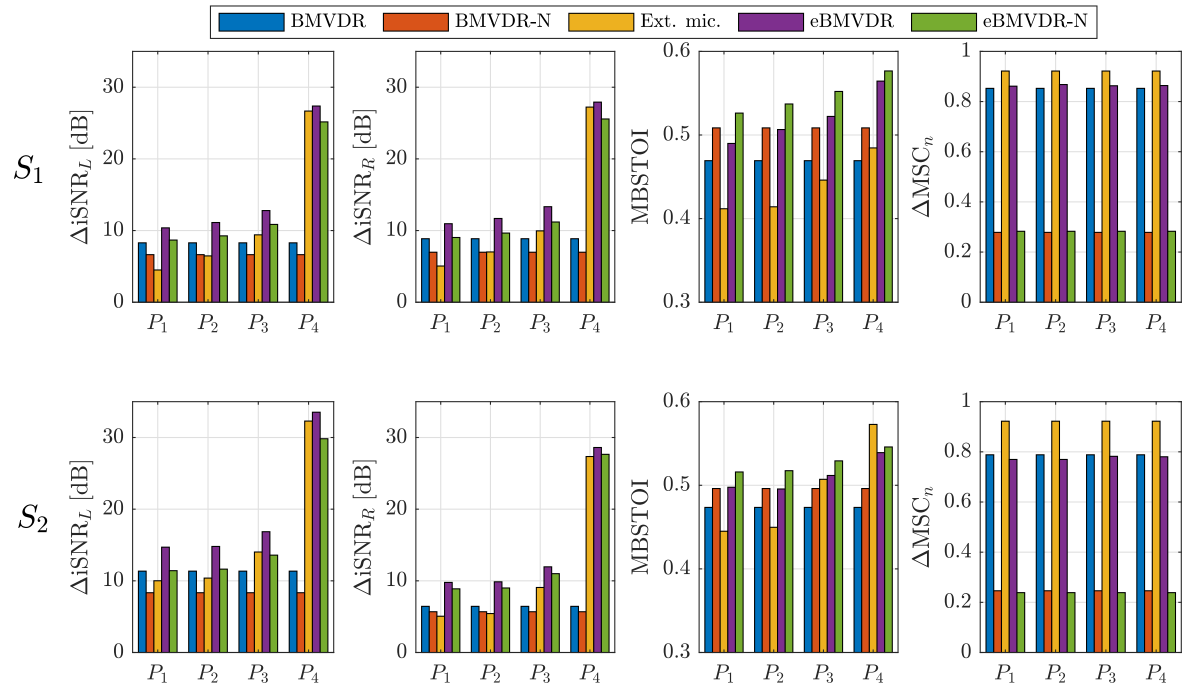


Fig. 11. Intelligibility-weighted SNR improvement $\Delta\mathrm{iSNR}_L$ and $\Delta\mathrm{iSNR}_R$, modified binaural short-time objective intelligibility (MBSTOI), and the MSC error $\Delta\mathrm{MSC}_n$ of the noise component for the BMVDR beamformer, the BMVDR-N beamformer, the external microphone signal, the eBMVDR beamformer and the eBMVDR-N beamformer, for the different positions of the external microphone. The results are shown for speaker S_1 (top row) and speaker S_2 (bottom row).

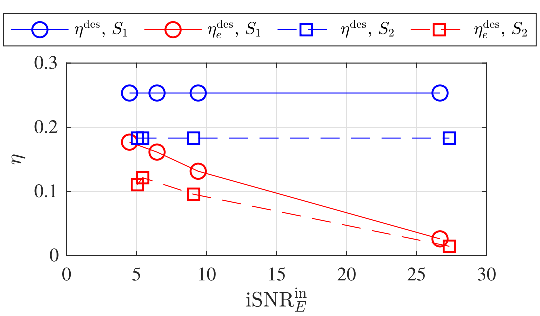


Fig. 12. Mixing parameters η_e^{des} and η^{des} (averaged over all frequencies) leading to the desired output MSC of the noise component for the different external microphone positions, mapped to the respective input iSNRs in the external microphone signal.

compared to the eBMVDR beamformer is different for the left and the right side but remains approximately constant for the different positions of the external microphone.

For both speakers S_1 and S_2, Fig. 12 depicts the mixing parameters η_e^{des} and η^{des} (averaged over all frequencies) of the eBMVDR-N beamformer and the BMVDR-N beamformer, which lead to the desired output MSC $\mathrm{MSC}_n^{\mathrm{des}}$ of the noise component. The mixing parameters are plotted as a function of the input iSNR in the external microphone signal $\mathrm{iSNR}_E^{\mathrm{in}}$ for the different external microphone positions. It can be observed that the mixing parameter is always smaller for the eBMVDR-N beamformer than for the BMVDR-N beamformer and decreases with increasing input iSNR in the external microphone signal. Further, the mixing parameter is always smaller for speaker S_2 than for speaker S_1, i.e., if the speaker is not positioned in front of the listener.

For speaker S_2 and the external microphone at position $P_{3,2}$, Fig. 13 depicts the spectrograms of the left reference microphone signal, the external microphone signal, and the left output signal of the BMVDR, BMVDR-N, eBMVDR and eBVMDR-N beamformers. The speech components in all signals were equalized in power. It can be observed that using or incorporating the external microphone (bottom row) enables the noise to be more reduced compared to not using or incorporating the external microphone (top row). Furthermore, it can be observed that due to the mixing less noise is reduced by the BMVDR-N and eBMVDR-N beamformers (right column) compared to the BMVDR and eBMVDR beamformers (mid column).

In conclusion, the experimental results in this section showed that for all considered positions of the external microphone and the speaker the iSNR improvement is larger for the eBMVDR-N beamformer (incorporating the external microphone) than for the BMVDR-N beamformer (using only the head-mounted microphones) and for the external microphone signal (except for P_4). In addition, the mixing parameter leading to the same desired output MSC of the noise component is always smaller for the eBMVDR-N beamformer than for the BMVDR-N beamformer. All experimental results in this section are in line with the theoretical findings of the previous sections. In future research we will investigate the incorporation of multiple external microphones, especially to efficiently estimate the RTF vectors of the desired source.

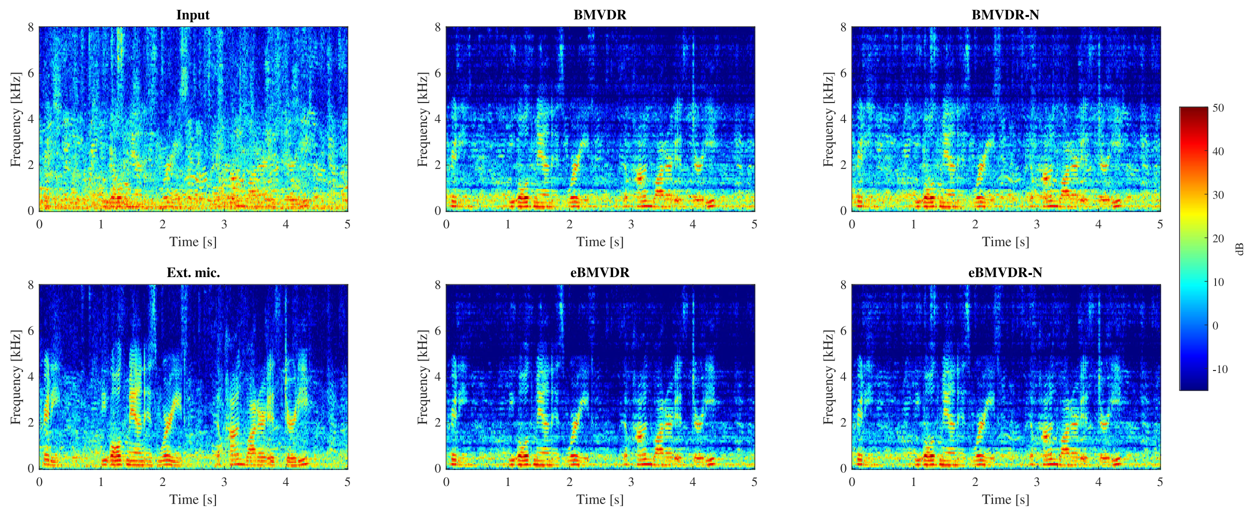


Fig. 13. Spectrograms of the left reference microphone signal (Input), the external microphone signal (Ext. mic.) and the left output signal of the BMVDR, BMVDR-N, eBMVDR and eBVMDR-N beamformers (speaker S_2 and external microphone at position $P_{3,2}$).

VI. Conclusion

In this paper, we analytically showed for an arbitrary noise field and without making any assumptions about the position of the desired source that by incorporating an external microphone in the BMVDR-N beamformer 1) a larger output SNR can be obtained for the same mixing parameter, 2) the same output SNR can be obtained for a larger mixing parameter, and 3) the same desired output MSC of the noise component can be obtained for a smaller mixing parameter. The obtained analytical expressions were first validated using simulated anechoic acoustic transfer functions. In addition, experimental results using recorded signals in a realistic reverberant environment showed that incorporating an external microphone in the BMVDR-N beamformer enables the output SNR to be significantly improved compared to using only the head-mounted microphone signals while preserving the spatial impression of the noise component.

References

[1] V. Hamacher, U. Kornagel, T. Lotter, and H. Puder, "Binaural signal processing in hearing aids: Technologies and algorithms," in *Proc. Adv. Digit. Speech Transmiss.*, 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 Speech Enhancement*. Hoboken, NJ, USA: Wiley, 2018, ch. 18, pp. 413–442.

[4] A. W. Bronkhorst and R. Plomp, "The effect of head-induced interaural time and level differences on speech intelligibility in noise," *J. Acoust. Soc. Am.*, vol. 83, no. 4, pp. 1508–1516, Apr. 1988.

[5] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA, USA: MIT Press, 1997.

[6] 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. Am.*, vol. 115, no. 2, pp. 833–843, Feb. 2004.

[7] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality," *J. Acoust. Soc. Am.*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.

[8] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears," *J. Acoust. Soc. Am.*, vol. 25, no. 5, pp. 975–979, Sep. 1953.

[9] 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.

[10] 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.

[11] 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/ACM Trans. Audio, Speech Lang. Process.*, vol. 23, no. 12, pp. 2384–2397, Dec. 2015.

[12] 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.

[13] 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.

[14] W. Pu, J. Xiao, T. Zhang, and Z.-Q. Luo, "A penalized inequality-constrained minimum variance beamformer with applications in hearing aids," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2017, pp. 175–179.

[15] 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.

[16] T. Klasen, 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.

[17] D. Marquardt, "Development and evaluation of psychoacoustically motivated binaural noise reduction and cue preservation techniques," Ph.D. dissertation, Dept. Med. Physics Acoust., Carl von Ossietzky Universität Oldenburg, 2015.

[18] 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. 1261–1274, Jul. 2018.

[19] J. Thiemann, M. Müller, D. Marquardt, S. Doclo, and S. van de Par, "Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene," *EURASIP J. Adv. Signal Process.*, vol. 2016, Feb. 2016, Art. no. 12.

[20] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, "Binaural speech enhancement with spatial cue preservation utilising simultaneous masking," in *Proc. Eur. Signal Process. Conf.*, Greece, Aug. 2017, pp. 628–632.

[21] 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.

[22] A. Bertrand and M. Moonen, “Robust distributed noise reduction in hearing aids with external acoustic sensor nodes,” *EURASIP J. Adv. Signal Process.*, vol. 2009, Jan. 2009, Art. no. 530435.

[23] J. Szurley, A. Bertrand, B. van Dijk, and M. Moonen, “Binaural noise cue preservation in a binaural noise reduction system with a remote microphone signal,” *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 24, no. 5, pp. 952–966, May 2016.

[24] D. Yee, H. Kamkar-Parsi, R. Martin, and H. Puder, “A noise reduction post-filter for binaurally-linked single-microphone hearing aids utilizing a nearby external microphone,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 26, no. 1, pp. 5–18, Jan. 2018.

[25] R. Ali, T. van Waterschoot, and M. Moonen, “Completing the RTF vector for an MVDR beamformer as applied to a local microphone array and an external microphone,” in *Proc. Int. Workshop Acoust. Signal Enhancement*, Tokyo, Japan, Sep. 2018, pp. 211–215.

[26] N. Gößling and S. Doclo, “Relative transfer function estimation exploiting spatially separated microphones in a diffuse noise field,” in *Proc. Int. Workshop Acoust. Signal Enhancement*, Tokyo, Japan, Sep. 2018, pp. 146–150.

[27] J. M. Kates, K. H. Arehart, R. K. Muralimanohar, and K. Sommerfeldt, “Externalization of remote microphone signals using a structural binaural model of the head and pinna,” *J. Acoust. Soc. Am.*, vol. 143, no. 5, pp. 2666–2677, May 2018.

[28] N. Gößling and S. Doclo, “RTF-steered binaural MVDR beamforming incorporating an external microphone for dynamic acoustic scenarios,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brighton, UK, May 2019, pp. 416–420.

[29] R. Ali, G. Bernardi, T. van Waterschoot, and M. Moonen, “Methods of extending a generalised sidelobe canceller with external microphones,” *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 27, no. 9, pp. 1349–1364, May 2019.

[30] J. M. Kates, K. H. Arehart, and L. O. Harvey Jr, “Integrating a remote microphone with hearing-aid processing,” *J. Acoust. Soc. Am.*, vol. 145, no. 6, pp. 3551–3566, Jun. 2019.

[31] N. Gößling, D. Marquardt, and S. Doclo, “Performance analysis of the extended binaural MVDR beamformer with partial noise estimation in a homogeneous noise field,” in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays*, San Francisco, USA, Mar. 2017, pp. 1–5.

[32] D. P. Jarret, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, “Rigid sphere room impulse response simulation: Algorithm and application,” *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1462–1472, Sep. 2012.

[33] W. S. Woods, E. Hadad, I. Merks, B. Xu, S. Gannot, and T. Zhang, “A real-world recording database for ad hoc microphone arrays,” in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2015, pp. 1–5.

[34] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[35] L. Wang and S. Doclo, “Correlation maximization based sampling rate offset estimation for distributed microphone arrays,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 3, pp. 571–582, Mar. 2016.

[36] D. Cherkassky and S. Gannot, “Blind synchronization in wireless acoustic sensor networks,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 3, pp. 651–661, Mar. 2017.

[37] J. Zhang, R. Heusdens, and R. C. Hendriks, “Rate-distributed spatial filtering based noise reduction in wireless acoustic sensor networks,” *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 26, no. 11, pp. 2015–2026, Jun. 2019.

[38] J. Zhang, R. Heusdens, and R. C. Hendriks, “Relative acoustic transfer function estimation in wireless acoustic sensor networks,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 10, pp. 1507–1519, Oct. 2019.

[39] 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.

[40] 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.

[41] M. Tohyama and A. Suzuki, “Interaural cross-correlation coefficients in stereo-reproduced sound fields,” *J. Acoust. Soc. Am.*, vol. 85, no. 2, pp. 780–786, Feb. 1989.

[42] A. Walther and C. Faller, “Interaural correlation discrimination from diffuse field reference correlations,” *J. Acoust. Soc. Am.*, vol. 133, no. 3, pp. 1496–1502, Mar. 2013.

[43] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, MA, USA: Cambridge Univ. Press, 2004.

[44] M. Jeub, M. Dörbecker, and P. Vary, “A semi-analytical model for the binaural coherence of noise fields,” *IEEE Signal Process. Lett.*, vol. 18, no. 3, pp. 197–200, Mar. 2011.

[45] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, “Intelligibility-weighted measures of speech-to-interference ratio and speech system performance,” *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[46] ANSI S3.5-1997, “Methods for calculation of the speech intelligibility index,” American National Standard (ANSI), New York, NY, USA, 1997.

[47] T. Gerkmann and R. C. Hendriks, “Unbiased MMSE-based noise power estimation with low complexity and low tracking delay,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 20, no. 4, pp. 1383–1393, May 2012.

[48] 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.

[49] A. Bertrand and M. Moonen, “Distributed node-specific LCMV beamforming in wireless sensor networks,” *IEEE Trans. Speech Audio Process.*, vol. 60, no. 1, pp. 233–246, Jan. 2012.

[50] R. Serizel, M. Moonen, B. van Dijk, and J. Wouters, “Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants,” *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.

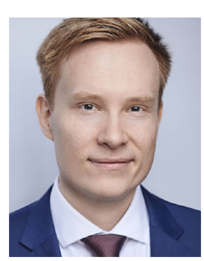
[51] S. Markovich-Golan and S. Gannot, “Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Brisbane, Australia, Apr. 2015, pp. 544–548.

[52] S. Markovich-Golan, S. Gannot, and W. Kellermann, “Performance analysis of the covariance-whitening and the covariance-subtraction methods for estimating the relative transfer function,” in *Proc. Eur. Signal Process. Conf.*, Rome, Italy, Sep. 2018, pp. 2499–2503.

[53] A. H. Andersen, J. M. de Haan, Z.-H. Tan, and J. Jensen, “Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions,” *Speech Commun.*, vol. 102, pp. 1–13, Sep. 2018.

[54] A. W. Bronkhorst, “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” *Acta Acustica United Acustica*, vol. 86, no. 1, pp. 117–128, Jan. 2000.

[55] N. Gößling, D. Marquardt, and S. Doclo, “Perceptual evaluation of binaural MVDR-based algorithms to preserve the interaural coherence of diffuse noise fields,” *Trends Hearing*, vol. 24, pp. 1–18, Apr. 2020.



Nico Gößling received the B.Sc. degree in media technology with focus on audio technology from the Hamburg University of Applied Sciences, Germany, in 2013 and the M.Sc. degree in hearing technology and audiology from the University of 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 include the area of signal processing for acoustical applications, more specifically microphone array processing for binaural hearing devices and acoustic sensor networks. In 2018, he was the recipient of 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.

Daniel Marquardt (Member, IEEE) received the Dipl.-Ing. in media technology from the Ilmenau University of Technology, Ilmenau, Germany, and the Dr.-Ing. degree in speech signal processing from the University of Oldenburg, Oldenburg, Germany, in 2010 and 2015, respectively. From 2015 to 2017, he was a Postdoctoral Researcher with the University of Oldenburg, Germany. Since 2017, he has been a Senior DSP Research Engineer with Starkey Hearing Technologies, Eden Prairie, MN, USA.

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, the 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.

Prof. Doclo was the recipient of the 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 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*.