Interaural Coherence Preservation for Binaural Noise Reduction Using Partial Noise Estimation and Spectral Postfiltering

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

***Abstract*—The objective of binaural speech enhancement algorithms is to reduce the undesired noise component, while preserving the desired speech source and the binaural cues of all sound sources. For the scenario of a single desired speech source in a diffuse noise field, an extension of the binaural multichannel Wiener filter (MWF), namely the MWF-IC, has been recently proposed, which aims to preserve the interaural coherence (IC) of the noise component. However, due to the large complexity of the MWF-IC, in this paper we propose several alternative algorithms at a lower computational complexity. First, we consider a quasi-distortionless version of the MWF-IC, denoted as minimum-variance-distortionless response (MVDR-IC). Second, we propose to preserve the IC of the noise component using the binaural MWF with partial noise estimation (MWF-N) and the binaural MVDR beamformer with partial noise estimation (MVDR-N), for which closed-form expressions exist. In addition, we show that for the MVDR-N a closed-form expression can be derived for the tradeoff parameter yielding a desired magnitude squared coherence (MSC) for the output noise component. Since contrary to the MWF-IC and the MWF-N the MVDR-IC and the MVDR-N do not take into account the spectro-temporal properties of the speech and the noise components, we propose to apply a spectral postfilter to the filter outputs, improving the noise reduction performance. The performance of all algorithms is compared in several diffuse noise scenarios. The simulation results show that both the MVDR-IC and the MVDR-N are able to preserve the MSC of the noise component, while generally the MVDR-IC shows a slightly better noise reduction performance at a larger complexity. Further, simulation results show that applying a spectral postfilter leads to a very similar performance for all considered algorithms in terms of noise reduction and speech distortion.**

***Index Terms*—Multi-channel Wiener filter, beamforming, hearing aids, binaural cues, noise reduction, interaural coherence.**

Manuscript received June 21, 2017; revised December 6, 2017 and March 4, 2018; accepted March 19, 2018. Date of publication April 4, 2018; date of current version April 24, 2018. This work was supported in part by the joint Lower Saxony-Israeli Project ATHENA and the Cluster of Excellence 1077 "Hearing4All," and in part by the German Research Foundation (DFG). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tan Lee. *(Corresponding author: Daniel Marquardt.)*

D. Marquardt and S. Doclo are with the Department of Medical Physics and Acoustics — Signal Processing Group University of Oldenburg, Oldenburg D-26111, Germany (e-mail: daniel.marquardt@uni-oldenburg.de; simon.doclo@uni-oldenburg.de).

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

Digital Object Identifier 10.1109/TASLP.2018.2823081

I. Introduction

NOISE reduction algorithms for head-mounted assistive listening devices (e.g., headsets, hearing aids, cochlear implants) are crucial to improve speech quality and intelligibility in background noise. For binaural microphone configurations, algorithms that use the microphone signals from the left and the right hearing device simultaneously are promising noise reduction techniques, because the spatial information captured by all microphones can be exploited [1]–[4]. Besides noise reduction and limiting speech distortion, another important objective of binaural algorithms is the preservation of the binaural cues of all sound sources. These binaural cues, i.e. the interaural level difference (ILD), the interaural time difference (ITD) and the interaural coherence (IC) are important for spatial awareness, i.e. for source localization and for determining the width of sound fields, and have a major impact on speech intelligibility due to the so-called binaural unmasking [5]–[8]. When the desired speech source is spatially separated from the interfering sources and background noise, a binaural hearing advantage compared to monaural hearing occurs. For example, in an anechoic environment with one desired speech source and one interfering source, both located in front of the listener, a speech reception threshold (SRT) corresponding to 50% speech intelligibility of about −8 dB is obtained [5]. If the sources are spatially separated, i.e., if the interfering source is not located in front of the listener, the SRT may decrease down to −20 dB, depending on the position of the interfering source. Although for reverberant environments this SRT difference is smaller than for anechoic environments, SRT differences for spatially separated sources up to 6 dB have been reported [9]. Furthermore, for scenarios with one desired speech source masked by a diffuse noise field, as considered in this paper, an improvement of the speech reception threshold (SRT) of 2–3 dB for both normal-hearing and hearing-impaired listeners has been reported [10], while no improvement in SRT can be observed if the desired speech source and the noise component are both coming from the same direction, i.e. contain the same spatial information [7]. For combined binaural noise reduction and cue preservation two main concepts have been established. In the first concept, the same real-valued spectro-temporal gain, which is typically derived from beamforming algorithms, blind source separation techniques or based on spatial assumptions, is applied to the reference microphone signals in both hearing devices [11]–[15]. Using this concept allows for perfect preservation of the instantaneous binaural cues, but may introduce audible artifacts.

In the second concept, all available microphone signals from both hearing devices are processed by different complex-valued filter vectors, combining spatial and spectral filtering [2], [4], [16]. Compared to the first concept, the second concept allows for more degrees of freedom to achieve increased noise reduction and less speech distortion but there is an unavoidable trade-off between noise reduction performance and binaural cue preservation. While these kind of algorithms typically preserve the binaural cues of the speech component, the binaural cues of the residual noise component are commonly distorted. Hence, a variety of algorithms have been proposed, aiming to also preserve the binaural cues of the residual noise component by adding additional terms or constraints, related to the preservation of the binaural cues, to the binaural noise reduction cost function [2], [16]–[26].

For the binaural multi-channel Wiener filter (MWF) [2] it has been theoretically proven in [16] that in the case of a single desired speech source the relative transfer function (RTF), comprising both the ILD and ITD cues, of the speech component is preserved, while the binaural cues of the residual noise component are distorted. More precisely, after applying the binaural MWF both output components exhibit the RTF of the speech component such that no spatial separation between the output speech component and the residual noise components exists and hence both components are perceived as coming from the same direction and the binaural hearing advantage can not be exploited by the auditory system.

In order to preserve the RTF of directional interfering sources, several extensions of the binaural MWF [16], [23], [27] and the binaural minimum-variance-distortionless-response (MVDR) beamformer [21], [24]–[26], [28], which is a special case of the binaural MWF, have been presented. Since these methods are either only applicable for directional interfering sources or not specifically designed for diffuse noise fields, in order to preserve the spatial characteristics of diffuse noise fields another extension of the binaural MWF, namely the MWF-IC, has been proposed in [22]. Since, contrary to directional interfering sources, the spatial characteristics of a diffuse noise field can not be properly described by the RTF but rather by the interaural coherence (IC), the MWF-IC aims at preserving the IC of the residual noise component, where the required amount of IC preservation can, e.g., be determined based on the IC discrimination ability of the human auditory system [22]. However, no closed-form solution for the MWF-IC exists such that one needs to resort to iterative numerical optimization methods, which are computationally intensive.

To reduce the computational complexity of the MWF-IC, in this paper we propose several alternative binaural noise reduction algorithms that aim to preserve the IC of the noise component, while achieving a psychoacoustically optimized trade-off between noise reduction and IC preservation.

Similarly as the binaural MVDR beamformer can be considered a distortionless version of the binaural MWF [2], we first present a quasi-distortionless version of the MWF-IC denoted as MVDR-IC, which does not depend on the time-varying spectral properties of the speech and the noise components. Secondly, we consider the binaural MWF with partial noise estimation (MWF-N) [16], [27] and its distortionless version, the binaural MVDR-N beamformer, for which closed form expressions exist. It has been shown in [16], [27] that the output signals of the binaural MWF-N consist of a mixture of the output signals of the binaural MWF with scaled noisy reference microphone signals, where a trade-off parameter determines the trade-off between noise reduction and binaural cue preservation of the noise component. For the MWF-N and the MVDR-N we derive analytical expressions for the IC and the magnitude squared coherence (MSC) of the output noise component. Furthermore, we show that a closed-form expression for the trade-off parameter yielding a desired MSC for the output noise component can be derived for the MVDR-N but not for the MWF-N. Although the MVDR-based algorithms (MVDR-IC, MVDR-N) are computationally advantageous compared to the MWF-based algorithms (MWF-IC, MWF-N) they do not take the spectro-temporal properties of the speech and the noise components into account, hence limiting the noise reduction performance. Therefore, we propose to apply a speech presence probability (SPP) based single-channel spectral postfilter to the output of the MVDR-based algorithms, mimicking the performance of the MWF-based algorithms at a much lower complexity.

The performance of the proposed algorithms is objectively evaluated in a reverberant cafeteria and office environment for one desired speech source at different spatial positions in a diffuse(-like) noise field. The simulation results show that both the MVDR-IC and the MVDR-N are able to preserve the MSC of the noise component, while generally the MVDR-IC shows a slightly better noise reduction performance. Further simulation results show that applying a spectral postfilter improves the noise reduction performance as indicated by the objective performance measure frequency-weighted segmental SNR.

The paper is structured as follows. In Section II-A, the configuration and notation of the considered binaural setup is described. In Section III, two state-of-the-art binaural noise reduction algorithms, namely the binaural MWF and the MWF-IC, are briefly reviewed and a quasi-distortionless version of the MWF-IC, denoted as MVDR-IC, is presented. In Section IV-A, the MWF-N and the MVDR-N are presented and analytical expressions for their noise reduction and speech distortion performance and for the output IC and MSC of the noise component are derived. Section V discusses how the trade-off parameter in the MVDR-IC, the MWF-N and the MVDR-N, yielding a desired MSC for the output noise component, can be determined. In Section VI, the application of a single-channel spectral postfilter on the output of the MVDR-based algorithms is presented. In Section VII, the performance of the proposed algorithms is evaluated.

II. Configuration and Notation

A. Microphone Signals and Output Signals

Consider the binaural configuration in Fig. 1 with $M = M_0 + M_1$ microphones, where M_0 and M_1 denote the number of microphones on the left and the right hearing device,

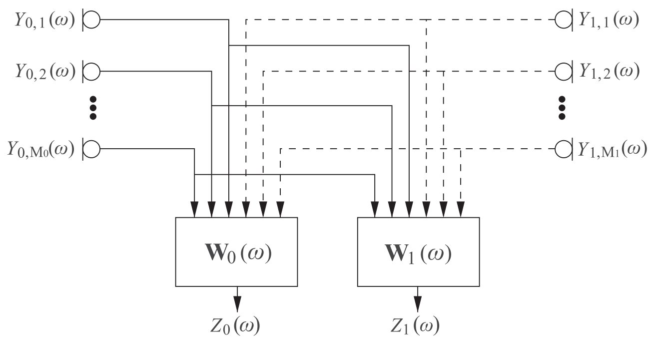

Fig. 1. Binaural hearing device configuration.

respectively. In the frequency-domain, the m-th microphone signal of the left hearing device $Y_{0,m}(\omega)$ can be written as

$$Y_{0,m}(\omega) = X_{0,m}(\omega) + V_{0,m}(\omega), \tag{1}$$

with ω the normalized radian frequency, $X_{0,m}(\omega)$ the speech component and $V_{0,m}(\omega)$ the diffuse noise component, respectively. The m-th microphone signal of the right hearing device $Y_{1,m}(\omega)$ is defined similarly. For the sake of readability, the frequency variable ω will be omitted in the remainder of the paper. The M-dimensional stacked signal vector $\mathbf{Y}$ is equal to

$$\mathbf{Y} = [Y_{0,1} \ldots Y_{0,M_0}\; Y_{1,1} \ldots Y_{1,M_1}]^T, \tag{2}$$

which can be written as

$$\mathbf{Y} = \mathbf{X} + \mathbf{V}, \tag{3}$$

where the vectors $\mathbf{X}$ and $\mathbf{V}$ are defined similarly as $\mathbf{Y}$. Assuming a desired speech source S, the speech component $\mathbf{X}$ can be written as

$$\mathbf{X} = S\mathbf{A}, \tag{4}$$

with $\mathbf{A}$ the acoustic transfer function (ATF) between the microphones and the speech source taking into account the head shadow effect and reverberation. For conciseness, the reference microphone signals $Y_{1,1}$ (first microphone in the left hearing device) and $Y_{1,1}$ (first microphone in the right hearing device) are denoted as Y_0 and Y_1, respectively, which can be written as

$$Y_0 = \mathbf{e}_0^T\mathbf{Y}, \quad Y_1 = \mathbf{e}_1^T\mathbf{Y}, \tag{5}$$

where $\mathbf{e}_0$ and $\mathbf{e}_1$ are M-dimensional selector vectors with $\mathbf{e}_0(1) = 1$ and $\mathbf{e}_1(M_0+1) = 1$, and all other elements equal to 0. The speech correlation matrix is defined as

$$\mathbf{R}_\mathrm{x} = \mathcal{E}\left\{\mathbf{X}\mathbf{X}^H\right\} = \Phi_\mathrm{s}\mathbf{A}\mathbf{A}^H, \tag{6}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator and $\Phi_\mathrm{s} = \mathcal{E}\left\{|S|^2\right\}$ denotes the power spectral density (PSD) of the speech source. The cross-correlation vectors between the speech component in all microphone signals and the speech component in the reference microphone signals are equal to

$$\mathbf{r}_{\mathrm{x},0} = \mathcal{E}\{\mathbf{X}X_0^*\} = \mathbf{R}_\mathrm{x}\mathbf{e}_0 = \Phi_\mathrm{s}\mathbf{A}A_0^*, \tag{7}$$

$$\mathbf{r}_{\mathrm{x},1} = \mathcal{E}\{\mathbf{X}X_1^*\} = \mathbf{R}_\mathrm{x}\mathbf{e}_1 = \Phi_\mathrm{s}\mathbf{A}A_1^*. \tag{8}$$

The PSD and the cross spectral density (CSD) of the speech component in the reference microphone signals are equal to

$$\Phi_{\mathrm{x},0} = \mathcal{E}\left\{|X_0|^2\right\} = \mathbf{e}_0^T\mathbf{R}_\mathrm{x}\mathbf{e}_0 = \Phi_\mathrm{s}|A_0|^2, \tag{9}$$

$$\Phi_{\mathrm{x},1} = \mathcal{E}\left\{|X_1|^2\right\} = \mathbf{e}_1^T\mathbf{R}_\mathrm{x}\mathbf{e}_1 = \Phi_\mathrm{s}|A_1|^2, \tag{10}$$

$$\Phi_{\mathrm{x},01} = \mathcal{E}\{X_0X_1^*\} = \mathbf{e}_0^T\mathbf{R}_\mathrm{x}\mathbf{e}_1 = \Phi_\mathrm{s}A_0A_1^*. \tag{11}$$

The noise correlation matrix is defined as

$$\mathbf{R}_\mathrm{v} = \mathcal{E}\left\{\mathbf{V}\mathbf{V}^H\right\}, \tag{12}$$

and the PSD and the CSD of the noise component in the reference microphone signals are equal to

$$\Phi_{\mathrm{v},0} = \mathcal{E}\left\{|V_0|^2\right\} = \mathbf{e}_0^T\mathbf{R}_\mathrm{v}\mathbf{e}_0, \tag{13}$$

$$\Phi_{\mathrm{v},1} = \mathcal{E}\left\{|V_1|^2\right\} = \mathbf{e}_1^T\mathbf{R}_\mathrm{v}\mathbf{e}_1, \tag{14}$$

$$\Phi_{\mathrm{v},01} = \mathcal{E}\{V_0V_1^*\} = \mathbf{e}_0^T\mathbf{R}_\mathrm{v}\mathbf{e}_1. \tag{15}$$

For a homogeneous diffuse noise field, the noise correlation matrix can be written as [29]

$$\mathbf{R}_\mathrm{v}^\mathrm{diff} = \mathcal{E}\left\{\mathbf{V}\mathbf{V}^H\right\} = \Phi_\mathrm{v}\mathbf{\Gamma}, \tag{16}$$

where Φ_v denotes the diffuse noise PSD which is the same in all microphone signals and $\mathbf{\Gamma}$ denotes the spatial coherence matrix of the diffuse noise field. This frequency-dependent coherence can be computed, e.g., using a geometric head model [30], using a modified sinc-function [31] or using measured anechoic ATFs [12] (cf. Section VII-A).

Assuming statistical independence, the correlation matrix of the microphone signals $\mathbf{R}_\mathrm{y}$ is equal to

$$\mathbf{R}_\mathrm{y} = \mathbf{R}_\mathrm{x} + \mathbf{R}_\mathrm{v}, \tag{17}$$

and the output signals are equal to

$$Z_0 = \mathbf{W}_0^H\mathbf{Y} = \mathbf{W}_0^H\mathbf{X} + \mathbf{W}_0^H\mathbf{V} = Z_{\mathrm{x}0} + Z_{\mathrm{v}0}, \tag{18}$$

$$Z_1 = \mathbf{W}_1^H\mathbf{Y} = \mathbf{W}_1^H\mathbf{X} + \mathbf{W}_1^H\mathbf{V} = Z_{\mathrm{x}1} + Z_{\mathrm{v}1}, \tag{19}$$

with $\mathbf{W}_0$ and $\mathbf{W}_1$ the $M-$dimensional filter vectors. The $2M$-dimensional stacked weight vector $\mathbf{W}$ is equal to

$$\mathbf{W} = \begin{bmatrix}\mathbf{W}_0\\ \mathbf{W}_1\end{bmatrix}. \tag{20}$$

B. Instrumental Performance Measures

In this section, we define instrumental performance measures which will be used for the analytical part of the paper. The narrowband speech distortion (SD) is defined as

$$SD_0 = \frac{\mathbf{e}_0^T\mathbf{R}_\mathrm{x}\mathbf{e}_0}{\mathbf{W}_0^H\mathbf{R}_\mathrm{x}\mathbf{W}_0}, \qquad SD_1 = \frac{\mathbf{e}_1^T\mathbf{R}_\mathrm{x}\mathbf{e}_1}{\mathbf{W}_1^H\mathbf{R}_\mathrm{x}\mathbf{W}_1}. \tag{21}$$

The narrowband input SNR in the reference microphone signals is defined as

$$SNR_0^\mathrm{in} = \frac{\mathbf{e}_0^T\mathbf{R}_\mathrm{x}\mathbf{e}_0}{\mathbf{e}_0^T\mathbf{R}_\mathrm{v}\mathbf{e}_0}, \qquad SNR_1^\mathrm{in} = \frac{\mathbf{e}_1^T\mathbf{R}_\mathrm{x}\mathbf{e}_1}{\mathbf{e}_1^T\mathbf{R}_\mathrm{v}\mathbf{e}_1}. \tag{22}$$

The narrowband output SNR in the output microphone signals of the left and the right hearing device is defined as

$$SNR_0^{\mathrm{out}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathrm{x}} \mathbf{W}_0}{\mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_0}, \quad SNR_1^{\mathrm{out}} = \frac{\mathbf{W}_1^H \mathbf{R}_{\mathrm{x}} \mathbf{W}_1}{\mathbf{W}_1^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1}. \tag{23}$$

The input RTFs are defined as

$$RTF_{\mathrm{x}}^{\mathrm{in}} = \frac{X_0}{X_1} = \frac{A_0}{A_1}, \qquad RTF_{\mathrm{v}}^{\mathrm{in}} = \frac{V_0}{V_1}, \tag{24}$$

and the output RTFs are defined as

$$RTF_{\mathrm{x}}^{\mathrm{out}} = \frac{\mathbf{W}_0^H \mathbf{A}}{\mathbf{W}_1^H \mathbf{A}}, \qquad RTF_{\mathrm{v}}^{\mathrm{out}} = \frac{\mathbf{W}_0^H \mathbf{V}}{\mathbf{W}_1^H \mathbf{V}}. \tag{25}$$

The input IC of the noise component is defined as the normalized cross-correlation between the reference microphone signals, i.e.,

$$IC_{\mathrm{v}}^{\mathrm{in}} = \frac{\mathbf{e}_0^T \mathbf{R}_{\mathrm{v}} \mathbf{e}_1}{\sqrt{(\mathbf{e}_0^T \mathbf{R}_{\mathrm{v}} \mathbf{e}_0)(\mathbf{e}_1^T \mathbf{R}_{\mathrm{v}} \mathbf{e}_1)}}, \tag{26}$$

and the output IC is defined as the normalized cross-correlation between the output noise components, i.e.,

$$IC_{\mathrm{v}}^{\mathrm{out}} = \frac{\mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_0)(\mathbf{W}_1^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1)}}. \tag{27}$$

The MSC is equal to

$$MSC = |IC|^2. \tag{28}$$

III. Binaural MWF and MVDR Beamformer with IC Preservation

In this section we first review the binaural MVDR beamformer [4], the binaural MWF [2] and the binaural MWF with IC preservation (MWF-IC) [22]. In addition, we propose a quasi-distortionless version of the MWF-IC, denoted as MVDR-IC.

A. Binaural MVDR Beamformer and Binaural MWF

The binaural MVDR beamformer [4], [32] aims to minimize the output PSD of the noise component in both hearing devices, while preserving the speech component in the reference microphone signals. The constrained optimization problem for the left and the right hearing device is equal to

$$\min_{\mathbf{W}_0} \mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_0 \quad \text{subject to} \quad \mathbf{W}_0^H \mathbf{H}_0 = 1, \tag{29}$$

$$\min_{\mathbf{W}_1} \mathbf{W}_1^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1 \quad \text{subject to} \quad \mathbf{W}_1^H \mathbf{H}_1 = 1, \tag{30}$$

with

$$\mathbf{H}_0 = \frac{\mathbf{A}}{A_0}, \qquad \mathbf{H}_1 = \frac{\mathbf{A}}{A_1}, \tag{31}$$

the RTF vectors of the speech source. The solution to the optimization problem in (29) and (30) is equal to [32]

$$\mathbf{W}_{\mathrm{MVDR},0} = \frac{\mathbf{R}_{\mathrm{v}}^{-1} \mathbf{H}_0}{\mathbf{H}_0^H \mathbf{R}_{\mathrm{v}}^{-1} \mathbf{H}_0}, \quad \mathbf{W}_{\mathrm{MVDR},1} = \frac{\mathbf{R}_{\mathrm{v}}^{-1} \mathbf{H}_1}{\mathbf{H}_1^H \mathbf{R}_{\mathrm{v}}^{-1} \mathbf{H}_1}. \tag{32}$$

The binaural MWF [2], [16] produces an MMSE estimate of the speech components X_0 and X_1 in the reference microphone signals of both hearing aids. The binaural (speech-distortion weighted) MWF cost function is defined as

$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_0^H \mathbf{V} \\ \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}, \tag{33}$$

where the weighting parameter $\mu \geq 0$ enables to trade-off between speech distortion and noise reduction. The cost function in (33) can further be written as

$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathbf{W}^H \mathbf{R} \mathbf{W} - \mathbf{W}^H \mathbf{r}_{\mathrm{x}} - \mathbf{r}_{\mathrm{x}}^H \mathbf{W} + \Phi_{\mathrm{x},0} + \Phi_{\mathrm{x},1}, \tag{34}$$

with

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{\mathrm{x}} + \mu \mathbf{R}_{\mathrm{v}} & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{R}_{\mathrm{x}} + \mu \mathbf{R}_{\mathrm{v}} \end{bmatrix}, \quad \mathbf{r}_{\mathrm{x}} = \begin{bmatrix} \mathbf{r}_{\mathrm{x},0} \\ \mathbf{r}_{\mathrm{x},1} \end{bmatrix}, \tag{35}$$

with $\mathbf{r}_{\mathrm{x},0}$ and $\mathbf{r}_{\mathrm{x},1}$ defined in (7) and (8). The filter minimizing $J_{\mathrm{MWF}}(\mathbf{W})$ in (34) is equal to [2]

$$\mathbf{W}_{\mathrm{MWF}} = \mathbf{R}^{-1} \mathbf{r}_{\mathrm{x}}, \tag{36}$$

such that the filter vectors for the left and the right hearing aid can be written as

$$\mathbf{W}_{\mathrm{MWF},0} = (\mathbf{R}_{\mathrm{x}} + \mu \mathbf{R}_{\mathrm{v}})^{-1} \mathbf{r}_{\mathrm{x},0}, \tag{37}$$

$$\mathbf{W}_{\mathrm{MWF},1} = (\mathbf{R}_{\mathrm{x}} + \mu \mathbf{R}_{\mathrm{v}})^{-1} \mathbf{r}_{\mathrm{x},1}. \tag{38}$$

As has been shown in [2], [16], for a single desired speech source the binaural MWF can be decomposed into a binaural MVDR beamformer and a single-channel spectro-temporal postfilter G applied to the output of the MVDR beamformer, i.e.,

$$\mathbf{W}_{\mathrm{MWF},0} = G_0 \mathbf{W}_{\mathrm{MVDR},0}, \quad \mathbf{W}_{\mathrm{MWF},1} = G_1 \mathbf{W}_{\mathrm{MVDR},1}, \tag{39}$$

with

$$G_0 = G_1 = \frac{\rho}{\mu + \rho}, \qquad \rho = \Phi_{\mathrm{s}} \mathbf{A}^H \mathbf{R}_{\mathrm{v}}^{-1} \mathbf{A}, \tag{40}$$

where ρ denotes the output SNR of the MVDR beamformer (cf. (46)). Consequently, for a single speech source, for which $\mathbf{R}_{\mathrm{x}} = \Phi_{\mathrm{x},0} \mathbf{H}_0 \mathbf{H}_0^H = \Phi_{\mathrm{x},1} \mathbf{H}_1 \mathbf{H}_1^H$, the binaural MVDR beamformer can be considered a special case of the binaural MWF in (37) and (38) for $\mu \rightarrow 0$, i.e.,

$$\mathbf{W}_{\mathrm{MVDR},0} = \lim_{\mu \rightarrow 0} (\mathbf{R}_{\mathrm{x}} + \mu \mathbf{R}_{\mathrm{v}})^{-1} \mathbf{r}_{\mathrm{x},0}, \tag{41}$$

$$\mathbf{W}_{\mathrm{MVDR},1} = \lim_{\mu \rightarrow 0} (\mathbf{R}_{\mathrm{x}} + \mu \mathbf{R}_{\mathrm{v}})^{-1} \mathbf{r}_{\mathrm{x},1}. \tag{42}$$

Please note that these expressions hold for any value of the PSDs $\Phi_{\mathrm{x},0}$ and $\Phi_{\mathrm{x},1}$. In addition, for a homogeneous diffuse noise field, cf. (16), the binaural MVDR beamformer in (32) is equal to

$$\mathbf{W}_{\mathrm{MVDR},0} = \frac{\mathbf{\Gamma}^{-1} \mathbf{H}_0}{\mathbf{H}_0^H \mathbf{\Gamma}^{-1} \mathbf{H}_0}, \quad \mathbf{W}_{\mathrm{MVDR},1} = \frac{\mathbf{\Gamma}^{-1} \mathbf{H}_1}{\mathbf{H}_1^H \mathbf{\Gamma}^{-1} \mathbf{H}_1}, \tag{43}$$

which is independent of the diffuse noise PSD Φ_{n}. By substituting (39) in (25), it can be shown that the output RTF of the speech component and the noise component are equal to the input RTF of the speech component [16], i.e.,

$$RTF_{\mathrm{x}}^{\mathrm{out}} = RTF_{\mathrm{v}}^{\mathrm{out}} = \frac{A_0}{A_1} = RTF_{\mathrm{x}}^{\mathrm{in}}, \tag{44}$$

implying that both output components are perceived as directional sources coming from the speech direction. By substituting (39) in (21), the speech distortion of the binaural MWF is equal to [16]

$$SD_{\mathrm{MWF},0} = SD_{\mathrm{MWF},1} = \frac{(\mu+\rho)^2}{\rho^2}, \tag{45}$$

which is always larger than or equal to 1. By substituting (39) in (23), the output SNR of the binaural MWF (and the binaural MVDR beamformer) is equal to [2]

$$SNR_{\mathrm{MWF},0}^{\mathrm{out}} = SNR_{\mathrm{MWF},1}^{\mathrm{out}} = \rho. \tag{46}$$

Furthermore, by substituting (39) in (27) and using (28), it can be shown that for the binaural MWF (and the binaural MVDR beamformer) the output IC of the speech and the noise component are the same and equal to [22]

$$IC_{\mathrm{x}}^{\mathrm{out}} = IC_{\mathrm{v}}^{\mathrm{out}} = e^{j\,\angle RTF_{\mathrm{x}}^{\mathrm{in}}}, \tag{47}$$

such that

$$MSC_{\mathrm{x}}^{\mathrm{out}} = MSC_{\mathrm{v}}^{\mathrm{out}} = 1. \tag{48}$$

Hence, for a diffuse noise field the binaural MWF (and the binaural MVDR beamformer) will not preserve the frequency-dependent IC/MSC of the noise component. Consequently, both the output speech and noise component will be perceived as directional sources from the same direction such that no binaural unmasking can be exploited by the auditory system and the perceived width of the diffuse noise field will not be present in the output noise component.

To implement the binaural MVDR beamformer in (32), an estimate of the RTF vectors $\mathbf{H}_0$ and $\mathbf{H}_1$ and the noise correlation matrix $\mathbf{R}_{\mathrm{v}}$ is required, for which several methods have been proposed [33]–[35]. However, since accurately estimating RTFs in noisy and reverberant environments is not a trivial task, as an alternative anechoic RTFs of the binaural setup $\bar{\mathbf{H}}_0(\theta)$ and $\bar{\mathbf{H}}_1(\theta)$, with θ denoting the (estimated) direction of arrival (DOA) of the desired speech source, can be used. The anechoic RTFs can be obtained using geometric head models or using measured impulse responses. In the case of a diffuse noise field, instead of the noise correlation matrix $\mathbf{R}_{\mathrm{v}}$, only the spatial coherence matrix $\mathbf{\Gamma}$ is required, cf. (43). As already mentioned, this coherence matrix can again be obtained using geometric head models or using measured impulse responses. Using anechoic RTFs and the spatial coherence matrix allows to calculate the filter vectors off-line, which can then be stored and applied to the microphone signals based on the estimated DOA. To implement the spectral postfilter in (40), an estimate of the output SNR of the binaural MVDR beamformer ρ is required, which can be obtained by combining a noise PSD estimator, e.g., [36], [37], with the decision-directed approach [36]. Hence, by implementing the binaural MWF in a diffuse noise field using the decomposition in (39) the correlation matrices $\mathbf{R}_{\mathrm{x}}$ and $\mathbf{R}_{\mathrm{v}}$, more in particular the input speech and noise PSDs in the reference microphone signals, do not need to be estimated. Using the decomposition of the binaural MWF in (39) instead of the direct form in (37) and (38) allows to decouple spatial filtering ($\mathbf{W}_{\mathrm{MVDR}}$) from spectro-temporal filtering (G), such that a-priori assumptions about the acoustic scenario and/or DOA estimates can be exploited for the spatial filter while spectral filtering can be individually controlled. Furthermore, since the spatial content of the acoustic scenario can be assumed to be more stationary then the spectro-temporal content of the speech signal, different smoothing parameters can be applied for estimating the spatial content (DOA or RTFs) and the spectral content (Φ_{x} and Φ_{n}). This is not possible if an estimate of the correlation matrices $\mathbf{R}_{\mathrm{x}}$ and $\mathbf{R}_{\mathrm{v}}$ is used, since in that case the spatial and spectro-temporal content is estimated simultaneously.

B. Binaural MWF with Interaural Coherence Preservation (MWF-IC)

Since the binaural MWF does not preserve the binaural cues of the noise component, in [22] the MWF has been extended with an additional term related to the preservation of the output IC of the noise component. The IC preservation term is equal to

$$J_{\mathrm{IC}}(\mathbf{W}) = \left| IC_{\mathrm{v}}^{\mathrm{out}}(\mathbf{W}) - IC_{\mathrm{v}}^{\mathrm{des}} \right|^2 \tag{49}$$

$$= \left| \frac{\mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1}{\sqrt{\left(\mathbf{W}_0^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_0\right)\left(\mathbf{W}_1^H \mathbf{R}_{\mathrm{v}} \mathbf{W}_1\right)}} - IC_{\mathrm{v}}^{\mathrm{des}} \right|^2, \tag{50}$$

where $IC_{\mathrm{v}}^{\mathrm{des}}$ represents the desired output IC, which can, e.g., be chosen to be equal to the input noise component IC in (26) or defined based on models or measurements [23], [30]. Adding this term to the cost function of the binaural MWF in (34), the cost function of the MWF-IC is equal to [22]

$$J_{\mathrm{MWF-IC}}(\mathbf{W}) = J_{\mathrm{MWF}}(\mathbf{W}) + \lambda\, J_{\mathrm{IC}}(\mathbf{W}). \tag{51}$$

Since no closed-form expression is available for the filter vector $\mathbf{W}_{\mathrm{MWF-IC}}$ minimizing the non-linear cost function in (51), an iterative numerical optimization method has been used [22]. Furthermore, since for the MWF-IC a trade-off between noise component IC preservation and output SNR exists (controlled by the trade-off parameter λ), in [22] the amount of IC preservation has been controlled based on the IC discrimination ability of the auditory system, cf. Section V. It should be noted that, contrary to the binaural MWF, the MWF-IC explicitly requires an estimate of the noise correlation matrices $\mathbf{R}_{\mathrm{x}}$ and $\mathbf{R}_{\mathrm{v}}$.

C. Quasi-Distortionless Version of the MWF-IC (MVDR-IC)

Similarly as the binaural MVDR can be considered a distortionless version of the binaural MWF (cf. Section III-A), in this section we present a quasi-distortionless version of the MWF-IC, denoted as MVDR-IC, to achieve noise reduction and IC preservation of the noise component. In Section III-A it has been shown that the binaural MVDR beamformer can be considered a special case of the binaural MWF for $\mu \rightarrow 0$, independently of the input speech PSDs and the diffuse noise PSD. Similarly, the MVDR-IC cost function is a special case of the MWF-IC cost function in (51) by setting the parameter μ to a very small value and setting the speech and noise correlation matrices equal to

$$\mathbf{R}_{\mathrm{x}} = \mathbf{H}_0 \mathbf{H}_0^H, \qquad \mathbf{R}_{\mathrm{v}} = \mathbf{\Gamma}, \tag{52}$$

such that no estimate of the speech and noise PSDs is required. The cost function for the MVDR-IC is hence equal to

$$J_{\mathrm{MVDR-IC}}(\mathbf{W}) = J_{\mathrm{MWF-QD}}(\mathbf{W}) + \lambda\, J_{\mathrm{IC-QD}}(\mathbf{W}), \tag{53}$$

with

$$J_{\mathrm{MWF-QD}}(\mathbf{W}) = \mathbf{W}^H \mathbf{R}_{\mathrm{QD}} \mathbf{W} - \mathbf{W}^H \mathbf{H}_0 - \mathbf{H}_0^H \mathbf{W}, \tag{54}$$

and

$$\mathbf{R}_{\mathrm{QD}} = \begin{bmatrix} \mathbf{H}_0 \mathbf{H}_0^H + \mu \mathbf{\Gamma} & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{H}_0 \mathbf{H}_0^H + \mu \mathbf{\Gamma} \end{bmatrix}, \tag{55}$$

the quasi-distortionless version of the MWF cost function and

$$J_{\mathrm{IC-QD}}(\mathbf{W}) = \left| \frac{\mathbf{W}_0^H \mathbf{\Gamma} \mathbf{W}_1}{\sqrt{\left(\mathbf{W}_0^H \mathbf{\Gamma} \mathbf{W}_0\right)\left(\mathbf{W}_1^H \mathbf{\Gamma} \mathbf{W}_1\right)}} - IC_{\mathrm{v}}^{\mathrm{des}} \right|^2. \tag{56}$$

The filter vector $\mathbf{W}_{\mathrm{MVDR-IC}}(\lambda)$ can then be obtained using the same numerical optimization method as for the MWF-IC. Note that distortionless filtering would only be obtained for $\mu = 0$. However, since $\mu = 0$ leads to numerical problems in the optimization routine, we use a very small value, i.e. $\mu = 10^{-5}$, such that the matrix $\mathbf{R}_{\mathrm{QD}}$ in (55) is invertible. Since for this value some hardly noticeable speech distortion occurs, we use the term quasi-distortionless. Contrary to the MWF-IC, the MVDR-IC does not require an estimate of the correlation matrices $\mathbf{R}_{\mathrm{x}}$ and $\mathbf{R}_{\mathrm{v}}$. As for the binaural MVDR (cf. Section III-A), instead of the RTF vector $\mathbf{H}_0$, the anechoic RTF vector $\bar{\mathbf{H}}_0(\theta)$ can be used leading to fixed filter vectors that can be computed off-line. However, contrary to the MWF-IC, the MVDR-IC does not take the time-varying spectral properties of the speech and the noise component into account which leads to a low amount of signal distortions but also to limited noise reduction capabilities compared to the MWF-IC. Hence, the application of a spectral postfilter at the output of the MVDR-IC will be discussed in Section VI.

IV. Binaural MWF and MVDR Beamformer with Partial Noise Estimation

In this section, the binaural MWF with partial noise estimation (MWF-N) [16], [27] is briefly reviewed. Compared to the MWF-IC, the MWF-N is a more general approach for preserving the binaural cues of the noise component since no specific IC preservation constraint is used. In addition, we propose the binaural MVDR beamformer with partial noise estimation (MVDR-N), which is a special case of the MWF-N. For both the MWF-N and the MVDR-N we derive analytical expressions for the IC and the MSC of the output noise component.

A. Binaural MWF with Partial Noise Estimation (MWF-N)

The MWF-N produces an MMSE estimate of the speech components X_0 and X_1 and a scaled version of the noise components V_0 and V_1 in the reference microphone signals [16], [27]. The cost function is defined as [16]

$$J_{\mathrm{MWF-N}}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \eta V_0 - \mathbf{W}_0^H \mathbf{V} \\ \eta V_1 - \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}, \tag{57}$$

where the real-valued parameter η, with $0 \le \eta \le 1$, enables to trade-off between output SNR and preservation of the binaural cues of the noise component. For the special case of $\eta = 0$, the cost function in (57) reduces to the cost function of the binaural MWF in (33). In [16] is has been shown that the filter vectors minimizing (57) are equal to

$$\mathbf{W}_{\mathrm{MWF-N},0} = (1-\eta)\, \mathbf{W}_{\mathrm{MWF},0} + \eta \mathbf{e}_0, \tag{58}$$

$$\mathbf{W}_{\mathrm{MWF-N},1} = (1-\eta)\, \mathbf{W}_{\mathrm{MWF},1} + \eta \mathbf{e}_1, \tag{59}$$

resulting in a summation of the binaural MWF output signals (weighted with 1-η) and the reference microphone signals (weighted with η). By substituting (58) and (59) in (25), it has been shown in [16] that the MWF-N preserves the binaural cues of the speech component for all values of the trade-off parameter η, i.e.,

$$RTF_{\mathrm{x}}^{\mathrm{out}} = \frac{(1-\eta)\frac{\rho}{(\mu+\rho)} A_0 + \eta A_0}{(1-\eta)\frac{\rho}{(\mu+\rho)} A_1 + \eta A_1} = \frac{A_0}{A_1} = RTF_{\mathrm{x}}^{\mathrm{in}}. \tag{60}$$

By substituting (58) and (59) in (21), the speech distortion of the MWF-N is equal to [16]

$$SD_{\mathrm{MWF-N},0} = SD_{\mathrm{MWF-N},1} = \left(\frac{\mu + \rho}{\eta \mu + \rho} \right)^2. \tag{61}$$

Comparing (61) with (45) implies that when $\eta > 0$ the binaural MWF always yields a larger speech distortion than the MWF-N. This can be intuitively explained by the fact that mixing the output speech component of the MWF with the input speech component of the reference microphone signals partially compensates the speech distortion introduced by the spectral postfilter. By substituting (58) and (59) in (23), it has been shown in [16] that the output SNR of the MWF-N is equal to

$$SNR_{\mathrm{MWF-N},0}^{\mathrm{out}} = \frac{\rho \xi}{\left(\xi + \eta^2 \left(\tilde{\rho}_0 - 1\right)\right)}, \tag{62}$$

$$SNR_{\mathrm{MWF-N},1}^{\mathrm{out}} = \frac{\rho \xi}{\left(\xi + \eta^2 \left(\tilde{\rho}_1 - 1\right)\right)}, \tag{63}$$

with

$$\xi = \left(\frac{\eta \mu + \rho}{\mu + \rho} \right)^2, \quad \tilde{\rho}_0 = \rho \frac{\Phi_{\mathrm{v},0}}{\Phi_{\mathrm{x},0}}, \qquad \tilde{\rho}_1 = \rho \frac{\Phi_{\mathrm{v},1}}{\Phi_{\mathrm{x},1}}, \tag{64}$$

where $\tilde{\rho}_0$ and $\tilde{\rho}_1$ denote the SNR improvement of the MWF in the left and the right hearing device, respectively. Since the SNR improvement of the MWF is always larger than or equal to 1 [38], (62) and (63) imply that the output SNR of the MWF-N is always smaller than or equal to the output SNR of the binaural

MWF, i.e.,

$$SNR^{\mathrm{out}}_{\mathrm{MWF-N},0} \leq SNR^{\mathrm{out}}_{\mathrm{MWF},0}, \tag{65}$$

$$SNR^{\mathrm{out}}_{\mathrm{MWF-N},1} \leq SNR^{\mathrm{out}}_{\mathrm{MWF},1}, \tag{66}$$

which can be intuitively explained by the mixing of the output signals of the MWF with the noisy reference microphone signals.

Since for diffuse noise we are mainly interested in the IC and the MSC, in Appendix A we derive the output IC of the noise component of the MWF-N by substituting (58) and (59) in (27), resulting in

$$IC^{\mathrm{out}}_{\mathrm{v}} = \frac{\psi \Phi_{\mathrm{x},01} + \eta^2 \Phi_{\mathrm{v},01}}{\sqrt{(\psi \Phi_{\mathrm{x},0} + \eta^2 \Phi_{\mathrm{v},0})(\psi \Phi_{\mathrm{x},1} + \eta^2 \Phi_{\mathrm{v},1})}}, \tag{67}$$

with

$$\psi = (1-\eta)^2 \frac{\rho}{(\mu+\rho)^2} + 2\eta(1-\eta)\frac{1}{(\mu+\rho)}. \tag{68}$$

From (67), the output MSC of the noise component can then be calculated as

$$MSC^{\mathrm{out}}_{\mathrm{v}} = \frac{|\psi \Phi_{\mathrm{x},01} + \eta^2 \Phi_{\mathrm{v},01}|^2}{(\psi \Phi_{\mathrm{x},0} + \eta^2 \Phi_{\mathrm{v},0})(\psi \Phi_{\mathrm{x},1} + \eta^2 \Phi_{\mathrm{v},1})}. \tag{69}$$

Hence, for the binaural MWF, i.e. setting $\eta = 0$, the output MSC of the noise component is equal to 1. On the other hand, for $\eta = 1$ the factor ψ in (68) is equal to 0, such that the output MSC of the noise component is equal to the input MSC of the noise component, but obviously no noise reduction is achieved. Similarly as for the MWF-IC and the MVDR-IC (cf. Section III-B and III-C), for the MWF-N a substantial trade-off between IC preservation of the noise component and noise reduction performance exists. Hence, in Section V we will present a procedure to determine the trade-off parameter η yielding a psychoacoustically optimized desired output MSC for the noise component.

B. Binaural MVDR Beamformer With Partial Noise Estimation (MVDR-N)

Similarly as the MWF-N, the optimization problem of the binaural MVDR beamformer can be modified such that, in addition to perfectly preserving the speech component, it aims at preserving a scaled version of the noise component in the reference microphone signals. The MVDR-N constrained optimization problem for the left and the right hearing device is given by

$$\min_{\mathbf{W}_0} \mathcal{E}\left\{ \left| \eta V_0 - \mathbf{W}_0^H \mathbf{V} \right|^2 \right\} \quad \text{subject to} \quad \mathbf{W}_0^H \mathbf{H}_0 = 1, \tag{70}$$

$$\min_{\mathbf{W}_1} \mathcal{E}\left\{ \left| \eta V_1 - \mathbf{W}_1^H \mathbf{V} \right|^2 \right\} \quad \text{subject to} \quad \mathbf{W}_1^H \mathbf{H}_1 = 1. \tag{71}$$

Similarly to (58) and (59), the filter vectors solving (70) and (71) are equal to

$$\mathbf{W}_{\mathrm{MVDR-N},0} = (1-\eta)\mathbf{W}_{\mathrm{MVDR},0} + \eta \mathbf{e}_0, \tag{72}$$

$$\mathbf{W}_{\mathrm{MVDR-N},1} = (1-\eta)\mathbf{W}_{\mathrm{MVDR},1} + \eta \mathbf{e}_1, \tag{73}$$

Hence, the MVDR-N can be considered a special case of the MWF-N in (58) and (59) for $\mu \rightarrow 0$,i.e., neglecting the spectral postfilter in the binaural MWF. Due to the distortionless constraint for the speech component in (70) and (71), the binaural cues of the speech component are perfectly preserved and the MVDR-N does not introduce any speech distortion, i.e.,

$$SD_{\mathrm{MVDR-N},0} = SD_{\mathrm{MVDR-N},1} = 1, \tag{74}$$

and the output SNR of the MVDR-N can be calculated by setting $\mu = 0$ in the output SNR of the MWF-N in (62) and (63), i.e.,

$$SNR^{\mathrm{out}}_{\mathrm{MVDR-N},0} = \frac{\rho}{[1 + \eta^2 (\tilde{\rho}_0 - 1)]}, \tag{75}$$

$$SNR^{\mathrm{out}}_{\mathrm{MVDR-N},1} = \frac{\rho}{[1 + \eta^2 (\tilde{\rho}_1 - 1)]}. \tag{76}$$

As for the binaural MWF, the output SNR of the MVDR-N is always smaller than or equal to the output SNR of the MVDR, i.e.,

$$SNR^{\mathrm{out}}_{\mathrm{MVDR-N},0} \leq SNR^{\mathrm{out}}_{\mathrm{MVDR},0}, \tag{77}$$

$$SNR^{\mathrm{out}}_{\mathrm{MVDR-N},1} \leq SNR^{\mathrm{out}}_{\mathrm{MVDR},1}. \tag{78}$$

Setting $\mu = 0$ in (68), the expression for ψ simplifies to

$$\psi_0 = \frac{1-\eta^2}{\rho}, \tag{79}$$

such that the output MSC of the noise component for the MVDR-N can now be calculated by substituting (79) in (69), i.e.,

$$MSC^{\mathrm{out}}_{\mathrm{v}} = \frac{\left| \frac{1-\eta^2}{\rho} \Phi_{\mathrm{x},01} + \eta^2 \Phi_{\mathrm{v},01} \right|^2}{\left(\frac{1-\eta^2}{\rho} \Phi_{\mathrm{x},0} + \eta^2 \Phi_{\mathrm{v},0} \right) \left(\frac{1-\eta^2}{\rho} \Phi_{\mathrm{x},1} + \eta^2 \Phi_{\mathrm{v},1} \right)}. \tag{80}$$

Comparing (80) with (69), it can be observed that the expression for the output MSC of the noise component significantly simplifies for the MVDR-N compared to the MWF-N.

V. Psychoacoustically Optimized Trade-Off Parameter for the MWF-IC, the MVDR-IC, the MWF-N and the MVDR-N

For all considered algorithms, the trade-off parameters λ (MWF-IC and MVDR-IC) and η (MWF-N and MVDR-N) enable to set a frequency-dependent trade off between noise reduction and IC preservation of the noise component. For the MWF-IC it has been proposed in [22] to determine the set of trade-off parameters Λ for which the output MSC of the noise component for the filter $\mathbf{W}_{\mathrm{MWF-IC}}(\lambda)$, minimizing the cost function in (51), satisfies

$$\Lambda = \left\{ \lambda | \gamma_{\min} \leq MSC^{\mathrm{out}}_{\mathrm{v}}(\mathbf{W}_{\mathrm{MWF-IC}}(\lambda)) \leq \gamma_{\max} \right\}, \tag{81}$$

where $\gamma_{\min}$ and $\gamma_{\max}$ denote lower and upper bounds for the MSC of the output noise component. The frequency-dependent constraint boundaries $\gamma_{\min}$ and $\gamma_{\max}$ have been defined based on the results of subjective listening experiments and are depicted in Fig. 2 (cf. [22] for further details). The trade-off parameter λ_{opt} is then determined in an iterative search such that the inequality constraint in (81) is satisfied [22]. Since the MVDR-IC is a special case of the MWF-IC, the same strategy can be used

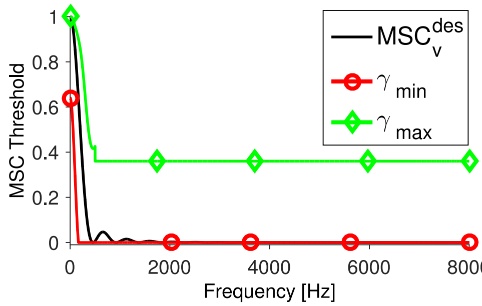

Fig. 2. Psychoacoustically motivated lower and upper MSC boundaries. For frequencies below 500 Hz, the boundaries depend on $MSC_{\mathrm{v}}^{\mathrm{des}}$. For frequencies above 500 Hz, the upper boundary is fixed to a value of 0.36 and the lower boundary is fixed to a value of 0" (cf. [22] for further details).

to determine the psychoacoustically optimized trade-off parameter λ_{opt} for the MVDR-IC. It should be realized that since the MWF-IC depends on the time-varying properties of the speech and the noise component (cf. Section III-B), the optimal trade-off parameter λ_{opt} for the MWF-IC needs to be determined for each time-frequency bin, resulting in a very computationally complex procedure. On the other hand, although an iterative search is also required to determine the optimal trade-off parameter λ_{opt} for the MVDR-IC, this only needs to be done once for each frequency bin, since the MVDR-IC corresponds to a fixed spatial filter (cf. Section III-C).

Similarly, for the MWF-N and the MVDR-N we would now like to determine the frequency-dependent trade-off parameter η_{opt}, yielding a desired magnitude squared coherence $MSC_{\mathrm{v}}^{\mathrm{des}}$ for the output noise component, where $MSC_{\mathrm{v}}^{\mathrm{des}}$ can be defined based on the MSC boundaries in Fig. 2. Setting the output MSC of the noise component for the MWF-N in (69) equal to the desired MSC and using (68) leads to

$$MSC_{\mathrm{v}}^{\mathrm{des}} = \frac{|\tilde{\psi}\Phi_{\mathrm{x},01} + \eta^2\,(\mu+\rho)^2\,\Phi_{\mathrm{v},01}|^2}{(\tilde{\psi}\Phi_{\mathrm{x},0} + \eta^2\,(\mu+\rho)^2\,\Phi_{\mathrm{v},0})(\tilde{\psi}\Phi_{\mathrm{x},1} + \eta^2\,(\mu+\rho)^2\,\Phi_{\mathrm{v},1})}. \tag{82}$$

with $\tilde{\psi} = \left(-\eta^2\,(2\mu+\rho) + 2\eta\mu + \rho\right)$. Solving this equation in closed-form requires the computation of the roots of a fourth order polynomial in the variable η, which is a rather cumbersome task. In the special case $A_0 = A_1$, corresponding to a speech source in front of the listener, it is possible to derive a closed-form expression for the trade-off parameter η_{opt} [39]. However, for the general case $A_0 \neq A_1$, considered in this paper, we will use an exhaustive search method for determining the trade-off parameter η_{opt} yielding a desired output MSC for the noise component. We first determine the set of possible trade-off parameters Υ for which the output MSC of the noise component for the resulting filter $\mathbf{W}_{\mathrm{MWF-N}}(\eta)$ in (58) and (59) satisfies the frequency-dependent boundaries $\gamma_{\min}$ and $\gamma_{\max}$, i.e.,

$$\Upsilon = \left\{\eta \,|\, \gamma_{\min} \leq MSC_{\mathrm{v}}^{\mathrm{out}}(\mathbf{W}_{\mathrm{MWF-N}}(\eta)) \leq \gamma_{\max}\right\}. \tag{83}$$

Since the output SNR of the MWF-N is monotonically decreasing with increasing η [16], the smallest value of the parameter η satisfying the MSC constraint in (83), i.e,

$$\eta_{\mathrm{opt}} = \min_{\eta\in\Upsilon}(\eta), \tag{84}$$

will result in the largest output SNR for both hearing devices. Again, it should be realized that since the MWF-N depends on the time-varying properties of the speech and the noise component (cf. Section IV-A), the optimal trade-off parameter η_{opt} needs to be determined for each time-frequency bin.

Since for the MVDR-N the expression for the output MSC of the noise component in (80) significantly simplifies compared to the MWF-N in (69), for the MVDR-N we can derive a closed form expression for the optimal trade-off parameter η_{opt} yielding a desired MSC for the output noise component. Setting $\mu = 0$ in (82) yields

$$MSC_{\mathrm{v}}^{\mathrm{des}} = \frac{\left|(1-\eta^2)\Phi_{\mathrm{x},01} + \eta^2\rho\Phi_{\mathrm{v},01}\right|^2}{((1-\eta^2)\Phi_{\mathrm{x},0} + \eta^2\rho\Phi_{\mathrm{v},0})\,((1-\eta^2)\Phi_{\mathrm{x},1} + \eta^2\rho\Phi_{\mathrm{v},1})}, \tag{85}$$

which corresponds to a second-order polynomial in the variable η^2

$$\left(1-\eta^2\right)^2 a + \eta^4 b + \eta^2\left(1-\eta^2\right)c, \tag{86}$$

with

$$a = MSC_{\mathrm{v}}^{\mathrm{des}}\Phi_{\mathrm{x},0}\Phi_{\mathrm{x},1} - |\Phi_{\mathrm{x},01}|, \tag{87}$$

$$b = \left(MSC_{\mathrm{v}}^{\mathrm{des}}\Phi_{\mathrm{v},0}\Phi_{\mathrm{v},1} - |\Phi_{\mathrm{v},01}|\right)\rho^2, \tag{88}$$

$$\begin{aligned} c = {} & MSC_{\mathrm{v}}^{\mathrm{des}}\left(\Phi_{\mathrm{x},1}\Phi_{\mathrm{v},0} + \Phi_{\mathrm{x},0}\Phi_{\mathrm{v},1}\right)\rho - \\ & \left(\Phi_{\mathrm{x},01}\Phi_{\mathrm{v},01}^* + \Phi_{\mathrm{x},01}^*\Phi_{\mathrm{v},01}\right)\rho. \end{aligned} \tag{89}$$

Solving for η yields the closed-form solution

$$\eta_{1,2} = \sqrt{\frac{-c + 2a \pm \sqrt{(c-2a)^2 - 4a(a+b-c)}}{2(a+b-c)}}, \tag{90}$$

where the smallest real-valued solution satisfying $0 \leq \eta \leq 1$ is the optimal solution. By using the assumption of a single desired speech source for which $\Phi_{\mathrm{x},0}\Phi_{\mathrm{x},1} = |\Phi_{\mathrm{x},01}|^2$, and by assuming a diffuse noise field, for which $\Phi_{\mathrm{n},0} = \Phi_{\mathrm{n},1}$, it can be shown that the parameters in (87)–(89) are equal to (after dividing each parameter with $\Phi_{\mathrm{x},0}\Phi_{\mathrm{x},1}$)

$$a = MSC_{\mathrm{v}}^{\mathrm{des}} - 1, \tag{91}$$

$$b = (MSC_{\mathrm{v}}^{\mathrm{des}} - MSC_{\mathrm{v}}^{\mathrm{in}})\tilde{\rho}_0\tilde{\rho}_1, \tag{92}$$

$$c = MSC_{\mathrm{v}}^{\mathrm{des}}\left(\tilde{\rho}_0 + \tilde{\rho}_1\right) - 2\Re\left\{\frac{A_0}{A_1}(IC_{\mathrm{v}}^{\mathrm{in}})^*\right\}\tilde{\rho}_0. \tag{93}$$

with $\tilde{\rho}_0 = \mathbf{H}_0^H\mathbf{\Gamma}^{-1}\mathbf{H}_0$ and $\tilde{\rho}_1 = \mathbf{H}_1^H\mathbf{\Gamma}^{-1}\mathbf{H}_1$ the SNR improvement of the MVDR beamformer defined in (64) and $IC_{\mathrm{v}}^{\mathrm{in}}$ and $MSC_{\mathrm{v}}^{\mathrm{in}}$ the input diffuse noise IC and MSC defined in (26) and (28). Hence, in the case of a diffuse noise field the optimal frequency-dependent parameter η_{opt} for the MVDR-N does not depend on the PSDs of the speech and the noise component, i.e., for a spatially stationary scenario this parameter is fixed for each frequency. Since for $\eta = 0$, $MSC_{\mathrm{v}}^{\mathrm{out}} = 1$ and for $\eta = 1$, $MSC_{\mathrm{v}}^{\mathrm{out}} = MSC_{\mathrm{v}}^{\mathrm{in}}$, for any $MSC_{\mathrm{v}}^{\mathrm{des}}$ with

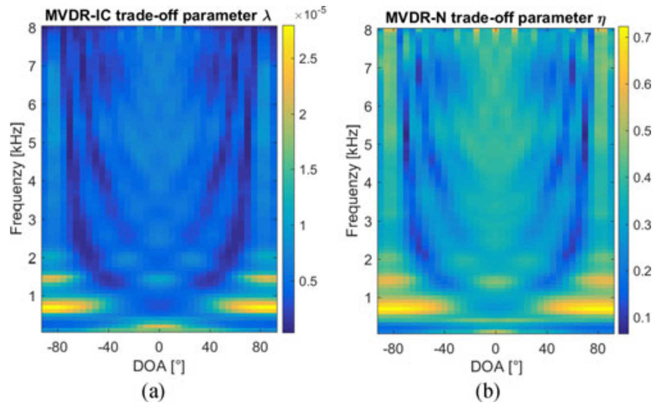

Fig. 3. Psychoacoustically optimized trade-off parameters $\lambda_{\rm opt}$ (MVDR-IC) and $\eta_{\rm opt}$ (MVDR-N) for different DOAs.

$MSC_{\rm v}^{\rm in} \leq MSC_{\rm v}^{\rm des} \leq 1$, a parameter $\eta_{\rm opt}$ can be found with $0 \leq \eta_{\rm opt} \leq 1$, for which the output SNR will lie between the output SNR of the MVDR beamformer and the input SNR. Based on these observations and the psychoacoustically motivated MSC boundaries in Fig. 2, it can be concluded that setting $MSC_{\rm v}^{\rm des} = \gamma_{\max}$ yields the optimal trade-off between a desired amount of MSC preservation and maximum noise reduction performance for the MVDR-N. The frequency-dependent optimal trade-off parameters $\lambda_{\rm opt}$ (MVDR-IC) and $\eta_{\rm opt}$ (MVDR-N) for different DOAs are depicted in Fig. 3.

VI. Spectral Postfilter for the MVDR-Based Algorithms

Although the psychoacoustically motivated MVDR-based binaural noise reduction and IC preservation algorithms (MVDR-IC, MVDR-N) are (quasi)-distortionless and are computationally advantageous compared to the MWF-based algorithms (MWF-IC, MWF-N), they do not take the time-varying spectral properties of the speech and the noise component into account, limiting their noise reduction performance. Therefore, we propose to apply a single-channel spectral filter at the output of the MVDR-IC and the MVDR-N. We propose to use a parametric Wiener filter, which can be computed for the left and the right hearing aid as

$$G_0 = \frac{\hat{SNR}_0^{\rm out}}{\mu + \hat{SNR}_0^{\rm out}}, \quad G_1 = \frac{\hat{SNR}_1^{\rm out}}{\mu + \hat{SNR}_1^{\rm out}}, \tag{94}$$

with $\hat{SNR}_0^{\rm out}$ and $\hat{SNR}_1^{\rm out}$ the SNR estimate of the left and the right output signal, respectively. We first estimate the PSD of the noise component in both signals using the single-channel SPP-based estimator described in [37] and then use these noise PSD estimates to estimate the SNRs using the decision-directed approach [36]. Since in general $\hat{SNR}_0^{\rm out} \neq \hat{SNR}_1^{\rm out}$, the gains G_0 and G_1 will not be equal and hence will distort the output ILD of the speech and the noise component. To avoid ILD distortions, a common gain for the left and the right hearing device is computed as

$$G = \sqrt{G_0\, G_1}. \tag{95}$$

TABLE I
Overview of the Considered Binaural Noise Reduction Algorithms

Algorithm	Filter vector	Trade-off parameter	Spectro-temporal filtering
MVDR	closed form, (32)	none	none
MWF	closed form, (39)	none	yes, output of MVDR
MWF-IC	numerical optimization, cf. Section III-B and [23]	λ, iterative search, cf. [23]	yes
MVDR-IC	numerical optimization, cf. Section III-B, III-C and [23]	λ, iterative search, cf. [23]	none
MVDR-ICP	numerical optimization, cf. Section III-B, III-C and [23]	λ, iterative search, cf. [23]	yes, output of MVDR-IC
MVDR-N	closed form, (72), (73)	η, closed form (90)	none
MVDR-NP	closed form, (72), (73), (97)	η, closed form (90)	yes, output of MVDR-N
MWF-N	closed form, (58), (59)	η, exhaustive search cf. Section V	yes, output of MVDR

Hence, for the hearing device with the better output SNR the common gain G will be lower than the corresponding Wiener gain, resulting in a increased noise reduction but also introducing more speech distortion, whereas for the hearing device with the lower output SNR the common gain G will be larger than the corresponding Wiener gain, resulting in a decreased noise reduction but also introducing less speech distortion. The MVDR-IC and the MVDR-N with spectral postfilter, denoted as MVDR-ICP and MVDR-NP, are then given by

$$\mathbf{W}_{\rm MVDR-ICP} = G\mathbf{W}_{\rm MVDR-IC}, \tag{96}$$

$$\mathbf{W}_{\rm MVDR-NP} = G\mathbf{W}_{\rm MVDR-N}. \tag{97}$$

It should be realized that the MVDR-ICP is not equivalent to the MWF-IC and the MVDR-NP is not equivalent to the MWF-N, although all mentioned algorithms apply spectro-temporal filtering. The main properties in terms of complexity for computing the filter vectors and the optimal trade-off parameter and exploitation of spectro-temporal properties are summarized in Table I.

In summary, the computational complexity of the MWF-IC is very high, since neither a closed-form expression for the filter vector nor for the optimal trade-off parameter exists. Furthermore, both the filter vector and the trade-off parameter need to be determined for each time-frequency bin. Although no closed-form expression for the MVDR-IC exist, the filter vectors and the trade-off parameter can be computed off-line. To exploit spectro-temporal properties, a postfilter can be applied to the MVDR-IC, resulting in the MVDR-ICP. Although the MWF-N can be easily implemented by mixing the output signals of the binaural MVDR beamformer with postfilter, i.e. the binaural MWF, with the noisy microphone signals, no closed form expression for the time-varying optimal trade-off parameter exists. Furthermore, as for the MWF-IC, both the filter vector and the trade-off parameter need to be determined for each time-frequency bin. As shown in Section IV-A, for the MVDR-N such a closed-form expression for the trade-off parameter exists. To exploit spectro-temporal properties, a postfilter can be applied to the MVDR-N, resulting in the MVDR-NP.

VII. Simulations

In this section, we compare the performance of the proposed algorithms in an office room and a cafeteria. We first compare the performance of the binaural MVDR beamformer, the MVDR-IC and the MVDR-N in terms of noise reduction performance and binaural cue preservation. Then, the impact of spectro-temporal postfiltering in the MWF, the MWF-N, the MVDR-ICP and the MVDR-NP on the noise reduction, speech distortion and binaural cue preservation performance is evaluated. Please note that we do not evaluate the performance of the MWF-IC due to its high computational complexity, which makes it infeasible for online applications.

A. Input Signals and Signal Statistics

To generate the microphone signals we used measured impulse responses for a binaural hearing aid setup in an office room ($T_{60} \approx 300$ ms) and a cafeteria ($T_{60} \approx 1250$ ms) [40]. Each hearing aid was equipped with 2 microphones, i.e. $M = 4$. For the office scenario the speech source was located either at -20°, 0° or 20°, while for the cafeteria scenario the speech source was located either at -35° or 0°. The reverberant speech component was generated by convolving the clean speech signals with the measured impulse responses taken from the Oldenburg Sentence Test database [41].

For the office scenario, a diffuse babble noise signal was generated using the method in [42], where the time-invariant spatial coherence matrix of the binaural setup was calculated using the ATFs of the anechoic impulse responses from [40], which were measured for angles ranging from -180° to 175° in steps of 5°. The (i, j)-th element of the spatial coherence matrix was calculated as

$$\mathbf{\Gamma}_{i,j} = \frac{\sum_{p=1}^{P} \bar{A}_i(\theta_p)\bar{A}_j^*(\theta_p)}{\sqrt{\sum_{p=1}^{P} |\bar{A}_i(\theta_p)|^2 \sum_{p=1}^{P} |\bar{A}_j(\theta_p)|^2}}, \tag{98}$$

with $\bar{A}(\theta_p)$ denoting the anechoic ATF at angle θ_p and $P = 72$ the total number of angles. For the cafeteria scenario, ambient noise including babble noise, clacking plates and interfering speakers, recorded in the same cafeteria, was used as noise component [40].

The overall length of the speech-and-noise signals was 20 s and the sampling frequency was equal to 16 kHz. The intelligibility-weighted input SNR (iSNR) [43] in the left reference microphone was set to 0 dB for each speech source position. The corresponding input iSNRs and frequency-weighted segmental SNRs (fwSegSNR) [44] in both reference microphones for each speech source position in the office room and the cafeteria are depicted in Table II.

We used the short-time Fourier transform (STFT) with frame length N overlapping by $N - L$ samples, e.g., for the left reference microphone

TABLE II
Intelligibility Weighted Input SNR and fwSegSNR for Each Speech Source Position in the Office Room and the Cafeteria

	Office			Cafeteria	
Angle	-20°	0°	20°	-35°	0°
$\mathrm{iSNR}_0^{\mathrm{in}}$ [dB]	0	0	0	0	0
$\mathrm{iSNR}_1^{\mathrm{in}}$ [dB]	-3.5	0.1	4.1	-8.1	0.1
$\mathrm{fwSegSNR}_0^{\mathrm{in}}$ [dB]	3	3.5	4	0.1	0.5
$\mathrm{fwSegSNR}_1^{\mathrm{in}}$ [dB]	1.7	3.5	5.5	-4.4	0

$$\begin{aligned} Y_0(k,i) &= \sum_{n=0}^{N-1} y_0(i\,L+n)\,w(n)\,e^{-j\Omega_k n}, \\ &= X_0(k,i) + V_0(k,i), \end{aligned} \tag{99}$$

with k the frequency index, i the frame segment index, $\Omega_k = 2\pi k/N$ the angular frequency, and $w(n)$ a square-root Hann window of length N. The segment length was set to $N = 512$ and L was set to 256.

B. Filter Vectors and Postfilter

For the MVDR, the MVDR-IC and the MVDR-N, the spatial coherence matrix $\mathbf{\Gamma}$ was calculated according to (98) and in (43) and (52) the anechoic RTF vectors $\bar{\mathbf{H}}_0(\theta_p)$ and $\bar{\mathbf{H}}_1(\theta_p)$ were used, assuming the DOA of the speech source to be known. For the MVDR-IC, the filter vector was calculated using an iterative numerical optimization method, and the frequency-dependent trade-off parameter λ_{opt} was calculated using an iterative search method. The filter vectors for the MVDR-N beamformer were calculated using the filter vectors of the binaural MVDR beamformer $\mathbf{W}_{\mathrm{MVDR},0}$ and $\mathbf{W}_{\mathrm{MVDR},1}$ according to (72) and (73), and the frequency-dependent trade-off parameter η_{opt} was calculated according to the closed-form expression in (90) using (91)–(93). The SNRs for the respective spectro-temporal postfilters in (39), (96) and (97) were estimated at the output of the MVDR, the MVDR-IC and the MVDR-N, respectively, using the SPP-based noise PSD estimator in combination with the decision-directed approach (cf. Section VI). For all postfilters the minimum gain was set to $G_{\min} = -10$ dB, the weighting parameter μ was set to 1 and the common gain was calculated according to (95). The MWF filter vectors were then used to calculate the MWF-N filter vectors in (58) and (59). The frequency-dependent trade-off parameter η_{opt} for each time-frequency bin was determined using the exhaustive search method described in Section V (cf. (83) and (84)), where 500 values for η, linearly spaced between 0 and 1, have been used.

C. Performance Measures

For the distortionless algorithms MVDR, MVDR-N and MVDR-IC, the noise reduction performance is evaluated using the intelligibility-weighted output SNR (iSNR) [43] in the left and the right hearing aid, which is defined as

$$iSNR^{\mathrm{out}} = \sum_{k=1}^{N-1} I(k)\,10\log_{10}(SNR^{\mathrm{out}}(k)), \tag{100}$$

where $I(k)$ is a weighting function [45]. $SNR^{\mathrm{out}}(k)$ is calculated according to (23), where the correlation matrices are calculated as

$$\mathbf{R}_{\mathrm{x}}(k) = \frac{1}{T}\sum_{i=0}^{T-1}\mathbf{X}(k,i)\mathbf{X}^H(k,i), \tag{101}$$

$$\mathbf{R}_{\mathrm{v}}(k) = \frac{1}{T}\sum_{i=0}^{T-1}\mathbf{V}(k,i)\mathbf{V}^H(k,i), \tag{102}$$

and T denotes the number of segments during the 20s of speech activity. The input SNR is similarly calculated according to (22). The better ear output SNR is defined as

$$iSNR_{\mathrm{be}}^{\mathrm{out}} = \max(iSNR_0^{\mathrm{out}}, iSNR_1^{\mathrm{out}}). \tag{103}$$

Since the purely energy-based SNR in (100) does not account for speech distortions introduced by spectro-temporal filtering, we also use the frequency-weighted segmental SNR (fwSegSNR) [44] as a combined measure for speech distortion and noise reduction. The better ear fwSegSNR is similarly defined as the better ear iSNR in (103).

The MSC preservation performance for the noise component is evaluated using the average broadband MSC error ΔMSC_{v} which is calculated as

$$\Delta MSC_{\mathrm{v}} = \frac{1}{N-1}\sum_{k=1}^{N-1}\left|MSC_{\mathrm{v}}^{\mathrm{in}}(k) - MSC_{\mathrm{v}}^{\mathrm{out}}(k)\right|, \tag{104}$$

where $MSC_{\mathrm{v}}^{\mathrm{in}}(k)$ and $MSC_{\mathrm{v}}^{\mathrm{out}}(k)$ are calculated as

$$MSC_{\mathrm{v}}^{\mathrm{in}}(k) = \frac{\left|\sum_i V_0(k,i)V_1^*(k,i)\right|^2}{\sum_i \left|V_0(k,i)\right|^2 \sum_i \left|V_1(k,i)\right|^2}, \tag{105}$$

$$MSC_{\mathrm{v}}^{\mathrm{out}}(k) = \frac{\left|\sum_i Z_{\mathrm{v}0}(k,i)Z_{\mathrm{v}1}^*(k,i)\right|^2}{\sum_i \left|Z_{\mathrm{v}0}(k,i)\right|^2 \sum_i \left|Z_{\mathrm{v}1}(k,i)\right|^2}. \tag{106}$$

To evaluate the binaural cue preservation of the speech component, we use a model of binaural auditory processing to calculate the histograms of the so-called reliable ILD and ITD cues, using Gammatone filters with a center frequency up to 4.8 kHz (ILD) and 1.3 kHz (ITD) [46]. For better comparison, the histograms for each Gammatone filter have been normalized to a height of 1.

D. Experimental Results

We first compare the iSNR and the noise component MSC preservation performance of the binaural MVDR, the MVDR-IC and the MVDR-N for different speech source positions.

The iSNR results for the office scenario are depicted in Fig. 4(a). As expected from the theoretical analysis, the output iSNR for the MVDR-IC and the MVDR-N is significantly lower than for the binaural MVDR, especially in the contralateral hearing aid, while the MVDR-IC outperforms the MVDR-N. The loss in better ear output iSNR compared to the binaural MVDR is between 0.3 and 0.5 dB for the MVDR-IC and between 0.3 and 1.1 dB for the MVDR-N.

For the cafeteria scenario, depicted in Fig. 4(b), the output iSNR of all algorithms is generally higher than for the office scenario. Again, compared to the MVDR, especially in the contralateral hearing aid, the output iSNR is lower for the MVDR-IC and the MVDR-N, where the loss in better ear output iSNR compared to the binaural MVDR is between 0.4 and 0.9 dB for the MVDR-IC and between 0.9 and 2.1 dB for the MVDR-N.

In terms of MSC error for the noise component (Fig. 5(d) (Office), Fig. 6(b) (Cafeteria)), the binaural MVDR results in a very large MSC error, which can be significantly reduced using the MVDR-IC and the MVDR-N, where for the cafeteria scenario the MSC error is generally lower. For each speech source position the MSC error is about the same, which implies that for both the MVDR-IC and the MVDR-N a suitable trade-off parameter yielding a desired MSC for the output noise component can be obtained. While the recorded ambient noise in the cafeteria is less spatially stationary than the artificially generated babble noise in the office room, these results hence show that also for a time-varying realistic noise field a controlled MSC preservation can be achieved using the proposed binaural cue preservation algorithms.

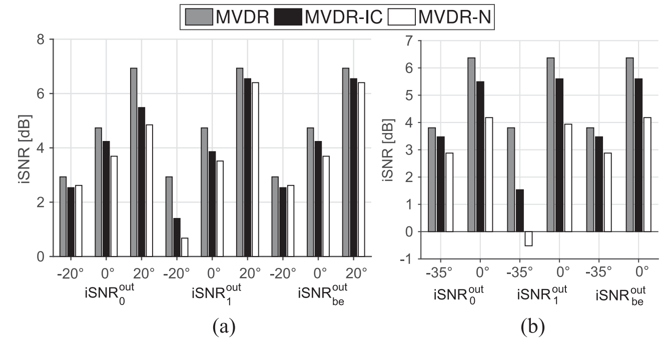

Fig. 4. Objective measures $\mathrm{iSNR}_0^{\mathrm{out}}$, $\mathrm{iSNR}_1^{\mathrm{out}}$, $\mathrm{iSNR}_{\mathrm{be}}^{\mathrm{out}}$ for the binaural MVDR beamformer, the MVDR-IC and the MVDR-N in (a) the office room and (b) the cafeteria for different positions of the speech source.

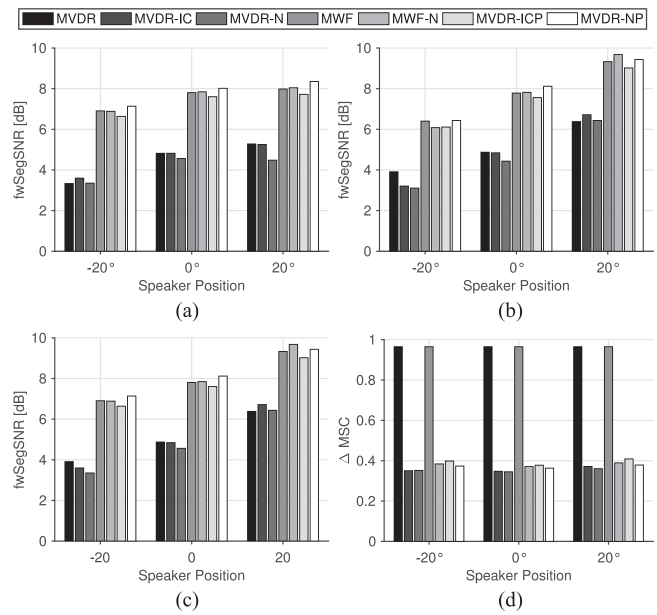

Fig. 5. Objective measures $\mathrm{fwSegSNR}_0^{\mathrm{out}}$, $\mathrm{fwSegSNR}_1^{\mathrm{out}}$, $\mathrm{fwSegSNR}_{\mathrm{be}}^{\mathrm{out}}$ and $\Delta\mathrm{MSC}_{\mathrm{v}}$ in the office room for different positions of the speech source.

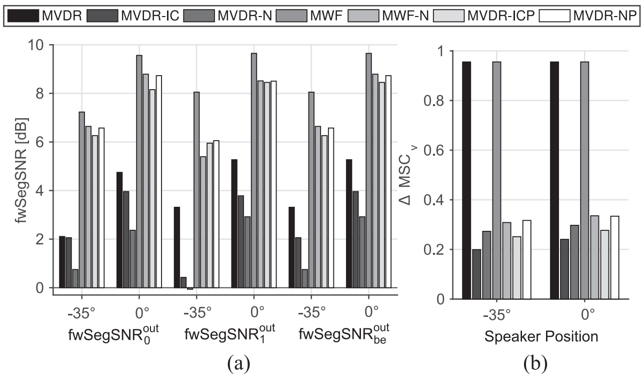

Fig. 6. Objective measures $\mathrm{fwSegSNR}_0^{\mathrm{out}}$, $\mathrm{fwSegSNR}_1^{\mathrm{out}}$, $\mathrm{fwSegSNR}_{\mathrm{be}}^{\mathrm{out}}$ and $\Delta\mathrm{MSC}_\mathrm{v}$ in the cafeteria for different positions of the speech source.

To evaluate the impact of the spectro-temporal postfilter in the binaural MWF, the MWF-N, the MVDR-ICP and the MVDR-NP the fwSegSNR results for the office and the cafeteria scenario for different speech source positions are depicted in Figs. 5 and 6, respectively.

For the office scenario, the distortionless algorithms MVDR, MVDR-IC and MVDR-N show rather similar results for the better ear fwSegSNR, whereas for the cafeteria scenario the MVDR shows the best performance. The spectro-temporal postfilter in the MWF, the MWF-N, the MVDR-ICP and the MVDR-NP leads to a significant improvement compared to the distortionless algorithms, while the MSC preservation algorithms MWF-N, MVDR-ICP and MVDR-NP show very similar results. Furthermore, both the MVDR-ICP and the MVDR-NP show very similar results as the MWF-N, while the computational complexity for the MVDR-ICP and the MVDR-NP is much lower compared to the MWF-N, where the trade-off parameter has to be computed for each time-frequency bin. As depicted in Figs. 5(d) and 6(b), spectral postfiltering has basically no impact on the MSC preservation capabilities of all algorithms. While the binaural MWF shows a very large noise component MSC error, the MSC error is significantly reduced for the MWF-N, the MVDR-ICP and the MVDR-NP. For a speech source at -20° in the office room, Fig. 7 depicts the histograms of the reliable ILD and ITD cues for the input speech component [Fig. 7(a)], the output speech component of the MVDR beamformer [Fig. 7(b)], the MVDR-IC [Fig. 7(c)] and the MVDR-N [Fig. 7(d)]. It should be noted that from the theoretical analysis in Section III, for the MVDR beamformer a perfect preservation of the binaural cues of the speech component is expected. However, from Fig. 7(a) and (b) it can be observed that this is not the case. This can be explained by the usage of the anechoic RTFs in the MVDR beamformer. Since the output binaural cues of the MVDR beamformer solely depend on A_0 and A_1 (cf. (44)), the binaural cues of the output speech component will be equal to the binaural cues of the corresponding anechoic signal from the same speech source position. Using the anechoic RTFs hence results in a decreased width of the histograms of the reliable ILD and ITD cues, which may change the overall impression of the perceived source width but not the perceived source position. Since for the MVDR-N a scaled version of the reference

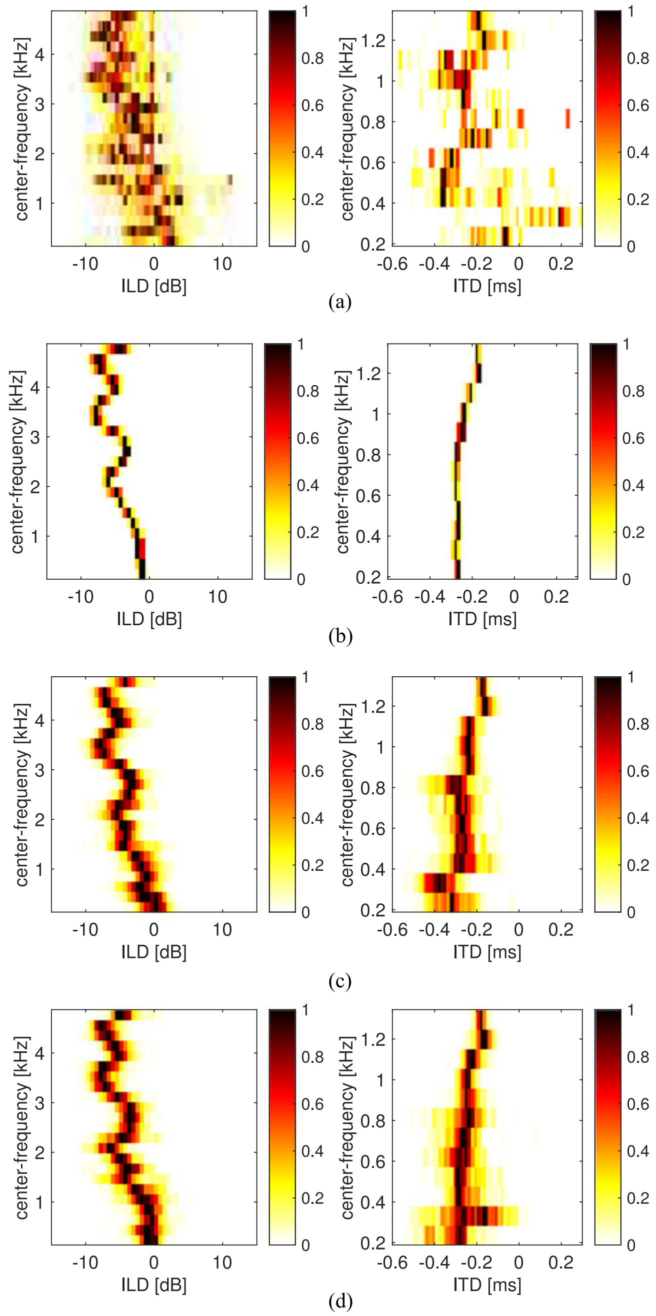

Fig. 7. Histograms of the reliable ILD and ITD cues for (a) the input speech component, and the output speech component of (b) the binaural MVDR, (c) the MVDR-IC and (d) the MVDR-N (speech source position of -20°).

microphone signals is added to the output of the MVDR beamformer, the histograms of the reliable ILD and ITD cues are wider for the MVDR-N compared to the MVDR beamformer, cf. Fig. 7(d). Furthermore, it can be noted that the output binaural cues of the MVDR-IC are very similar to the output binaural cues of the MVDR-N. Similar results are obtained for the other speech source positions in the office room and the cafeteria.

VIII. Conclusion

In this paper, we proposed several MVDR-based and MWF-based algorithms to preserve the binaural cues in diffuse noise scenarios. For all algorithms, the amount of IC preservation for the noise component was determined based on psychoacoustically motivated MSC boundaries. The main differences between the algorithms is their computational complexity to compute the filter vectors and the optimal trade-off parameter and whether they take the spectro-temporal properties of the speech and the noise component into account. For the MVDR-based algorithms, we have proposed a (quasi)-distortionless version of the MWF-IC, namely the MVDR-IC. In addition, for the MVDR-N we have derived a closed-form expression for the trade-off parameter yielding a desired MSC for the output noise component. Using simulations, we have shown that the MVDR-N and the MVDR-IC yield a very similar performance in terms of IC preservation, where the MVDR-IC shows a slightly better noise reduction performance. In addition, we have proposed to apply a spectro-temporal postfilter to the MVDR-IC and the MVDR-N. Simulation results show that the MWF-N, the MVDR-ICP and the MVDR-NP show a very similar performance in terms of noise reduction and binaural cue preservation. However, the computational complexity for the MVDR-ICP and the MVDR-NP is much lower compared to the MWF-N. For the MVDR-NP, the optimal trade-off parameter can be calculated using a closed-form expression and is fixed over time, whereas for the MWF-N, the trade-off parameter has to be calculated for each time-frequency bin using an exhaustive search method.

APPENDIX A
OUTPUT IC OF THE MWF-N

The output CSD and PSDs of the noise component for the MWF-N are equal to

$$\begin{aligned}\mathcal{E}\{Z_{\mathrm{v}0}Z_{\mathrm{v}1}^*\} &= \mathbf{W}_{\mathrm{MWF-N},0}^H\mathbf{R}_{\mathrm{v}}\mathbf{W}_{\mathrm{MWF-N},1}\\ &= (1-\eta)^2\mathbf{W}_{\mathrm{MWF},0}^H\mathbf{R}_{\mathrm{v}}\mathbf{W}_{\mathrm{MWF},1}+\eta^2\Phi_{\mathrm{v},01}\\ &\quad+2\eta(1-\eta)\big(\mathbf{W}_{\mathrm{MWF},0}^H\mathbf{R}_{\mathrm{v}}\mathbf{e}_1+\mathbf{e}_0^T\mathbf{R}_{\mathrm{v}}\mathbf{W}_{\mathrm{MWF},1}\big),\end{aligned} \tag{107}$$

$$\begin{aligned}\mathcal{E}\{|Z_{\mathrm{v}0}|^2\} &= \mathbf{W}_{\mathrm{MWF-N},0}^H\mathbf{R}_{\mathrm{v}}\mathbf{W}_{\mathrm{MWF-N},0}\\ &= (1-\eta)^2\mathbf{W}_{\mathrm{MWF},0}^H\mathbf{R}_{\mathrm{v}}\mathbf{W}_{\mathrm{MWF},0}+\eta^2\Phi_{\mathrm{v},0}\\ &\quad+2\eta(1-\eta)\Re\left\{\mathbf{W}_{\mathrm{MWF},0}^H\mathbf{R}_{\mathrm{v}}\mathbf{e}_0\right\},\end{aligned} \tag{108}$$

$$\begin{aligned}\mathcal{E}\{|Z_{\mathrm{v}1}|^2\} &= \mathbf{W}_{\mathrm{MWF-N},1}^H\mathbf{R}_{\mathrm{v}}\mathbf{W}_{\mathrm{MWF-N},1}\\ &= (1-\eta)^2\mathbf{W}_{\mathrm{MWF},1}^H\mathbf{R}_{\mathrm{v}}\mathbf{W}_{\mathrm{MWF},1}+\eta^2\Phi_{\mathrm{v},1}\\ &\quad+2\eta(1-\eta)\Re\left\{\mathbf{W}_{\mathrm{MWF},1}^H\mathbf{R}_{\mathrm{v}}\mathbf{e}_1\right\}.\end{aligned} \tag{109}$$

Using (39), the output CSD of the noise component in (107) is equal to

$$\begin{aligned}\mathcal{E}\{Z_{\mathrm{v}0}Z_{\mathrm{v}1}^*\} &= (1-\eta)^2\Phi_{\mathrm{x},01}\frac{\rho}{(\mu+\rho)^2}+\eta^2\Phi_{\mathrm{v},01}\\ &\quad+2\eta(1-\eta)\frac{1}{\mu+\rho}\Phi_{\mathrm{x},01}\\ &= \psi\Phi_{\mathrm{x},01}+\eta^2\Phi_{\mathrm{v},01},\end{aligned} \tag{110}$$

with ψ defined in (68). Calculating the output PSDs of the noise component in (108) and (109) in a similar way, the output IC of the noise component is equal to

$$IC_{\mathrm{v}}^{\mathrm{out}} = \frac{\psi\Phi_{\mathrm{x},01}+\eta^2\Phi_{\mathrm{v},01}}{\sqrt{(\psi\Phi_{\mathrm{x},0}+\eta^2\Phi_{\mathrm{v},0})(\psi\Phi_{\mathrm{x},1}+\eta^2\Phi_{\mathrm{v},1})}}. \tag{111}$$

REFERENCES

[1] V. Hamacher, U. Kornagel, T. Lotter, and H. Puder, "Binaural signal processing in hearing aids: Technologies and algorithms," in *Advances in Digital Speech Transmission*, New York, NY, USA: Wiley, 2008, pp. 401–429.

[2] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, Hoboken, NJ, USA: Wiley, 2010, pp. 269–302.

[3] J. Wouters, S. Doclo, R. Koning, and T. Francart, "Sound processing for better coding of monaural and binaural cues in auditory prostheses," *Proc. IEEE*, vol. 101, no. 9, pp. 1986–1997, Sep. 2013.

[4] S. Doclo, W. Kellermann, S. Makino, and S. 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.

[5] 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, 1988.

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

[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] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *J. Acoustical Soc. Amer.*, vol. 120, no. 1, pp. 331–342, 2006.

[9] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acta Acust. united with Acust.*, vol. 86, no. 1, pp. 117–128, Jan. 2000.

[10] I. Arweiler and J. M. Buchholz, "The influence of spectral characteristics of early reflections on speech intelligibility," *J. Acoust. Soc. Amer.*, vol. 130, no. 2, pp. 996–1005, 2011.

[11] T. Wittkop and V. Hohmann, "Strategy-selective noise reduction for binaural digital hearing aids," *Speech Commun.*, vol. 39, no. 1–2, pp. 111–138, Jan. 2003.

[12] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 1–14, 2006.

[13] K. Reindl, Y. Zheng, and W. Kellermann, "Analysis of two generic Wiener filtering concepts for binaural speech enhancement in hearing aids," in *Proc. 18th Eur. Signal Process. Conf.*, Aalborg, Denmark, Aug. 2010, pp. 989–993.

[14] K. Reindl *et al.*, "A stereophonic acoustic signal extraction scheme for noisy and reverberant environments," *Comput. Speech Lang.*, vol. 27, no. 3, pp. 726 – 745, 2013.

[15] R. Baumgärtel *et al.*, "Comparing binaural signal processing strategies I: Instrumental evaluation." *Trends Hearing*, vol. 19, pp. 1–16, 2015.

[16] 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.

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

[18] 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.,* Honolulu, HI, USA, Apr. 2007, pp. 5–8.

[19] T. Van den Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Honolulu, HI, USA, Apr. 2007, pp. 565–568.

[20] D. Marquardt, V. Hohmann, and S. Doclo, "Binaural cue preservation for hearing aids using multi-channel Wiener filter with instantaneous ITF preservation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Kyoto, Japan, Mar. 2012, pp. 21–24.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] E. Hadad *et al.*, "Comparison of two binaural beamforming approaches for hearing aids," in *Proc. IEEE Int. Conf. Acoust, Speech, Signal Process.*, New Orleans, LA, USA, Mar. 2017, pp. 236–240.

[27] 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.

[28] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, "Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation," in *Proc. 14th Int. Workshop Acoust. Signal Enhancement*, Juan-les-Pins, France, Sep. 2014, pp. 288–292.

[29] H. Cox, "Spatial correlation in arbitrary noise fields with application to ambient sea noise," *J. Acoust. Soc. Amer.*, vol. 54, no. 5, pp. 1289–1301, 1973.

[30] 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.

[31] I. Lindevald and A. Benade, "Two-ear correlation in the statistical sound fields of rooms," *J. Acoust. Soc. Amer.*, vol. 80, no. 2, pp. 661–664, Aug. 1986.

[32] 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.

[33] 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.

[34] 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, QLD, Australia, Apr. 2015, pp. 544–548.

[35] R. C. Hendriks and T. Gerkmann, "Noise correlation matrix estimation for multi-microphone speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 223–233, Jan. 2012.

[36] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust, Speech, Signal Process.*, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[37] T. Gerkmann and R. 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.

[38] S. Doclo and M. Moonen, "On the output SNR of the speech-distortion weighted multichannel Wiener filter," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 809–811, Dec. 2005.

[39] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural coherence preservation in MWF-based binaural noise reduction algorithms using partial noise estimation," in *Proc. IEEE Int. Conf. Acoust, Speech, Signal Process.*, Brisbane, QLD, Australia, Apr. 2015, pp. 654–658.

[40] H. Kayser, S. Ewert, J. Annemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *Eurasip J. Advances Signal Process.*, vol. 2009, 2009, Art no. 298605.

[41] K. Wagener, V. Kühnel, and B. Kollmeier, "Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test)," *Zeitschrift für Audiologie/Audiological Acoust.*, vol. 38, pp. 4–15, 1999.

[42] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *J. Acoust. Soc. Amer.*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.

[43] 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. Amer*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[44] H. Yi and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.

[45] ANSI S3.5-1997, *Methods for calculation of the speech intelligibility index*, American National Standard (ANSI), 1997.

[46] M. Dietz, S. D. Ewert, and V. Hohmann, "Auditory model based direction estimation of concurrent speakers from binaural signals," *Speech Commun.*, vol. 53, pp. 592–605, 2011.

Daniel Marquardt (S'12–M'15) received the Dipl.-Ing. degree 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 (S'95–M'03–SM'13) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, 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), Hamilton, ON, Canada. From 2007 to 2009, he was a Principal Scientist with NXP Semiconductors, Sound and Acoustics Group, Leuven, Belgium. Since 2009, he has been a Full Professor with the University of Oldenburg, Oldenburg Germany, and a Scientific Advisor with the Project Group Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany. His research interests include center around signal processing for acoustical and biomedical applications, more specifically microphone array processing, active noise control, acoustic sensor networks, and hearing aid processing. He was the recipient of the Master Thesis Award of the Royal Flemish Society of Engineers in 1997 (with Erik De Clippel), the Best Student Paper Award of the International Workshop on Acoustic Echo and Noise Control in 2001, the *EURASIP Journal on Advances in Signal Processing* Best Paper Award in 2003 (with Marc Moonen), and the IEEE Signal Processing Society Best Paper Award in 2008 (with Jingdong Chen, Jacob Benesty, and Arden Huang). He is member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing, the EURASIP Special Area Team on Acoustic, Speech and Music Signal Processing and the EAA Technical Committee on Audio Signal Processing. He was a Guest Editor for several special issues (IEEE Signal Processing Magazine, Elsevier Signal Processing) and an Associate Editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and *EURASIP Journal on Advances in Signal Processing*.