Interaural Coherence Preservation in Multi-Channel Wiener Filtering-Based Noise Reduction for Binaural Hearing Aids

Daniel Marquardt, *Student Member, IEEE*, Volker Hohmann, and Simon Doclo, *Senior Member, IEEE*

***Abstract*—Besides noise reduction an important objective of binaural speech enhancement algorithms is the preservation of the binaural cues of all sound sources. To this end, an extension of the binaural multi-channel Wiener filter (MWF), namely the MWF-ITF, has been proposed, which aims to preserve the Interaural Transfer Function (ITF) of the noise sources. However, the MWF-ITF is well-suited only for directional noise sources but not for, e.g., spatially isotropic noise, whose spatial characteristics cannot be properly described by the ITF but rather by the Interaural Coherence (IC). Hence, another extension of the binaural MWF, namely the MWF-IC, has been recently proposed, which aims to preserve the IC of the noise component. Since for the MWF-IC a substantial tradeoff between noise reduction and IC preservation exists, in this paper we propose a perceptually constrained version of the MWF-IC, where the amount of IC preservation is controlled based on the IC discrimination ability of the human auditory system. In addition, a theoretical analysis of the binaural cue preservation capabilities of the binaural MWF and the MWF-ITF for spatially isotropic noise fields is provided. Several simulations in diffuse noise scenarios show that the perceptually constrained MWF-IC yields a controllable preservation of the IC without significantly degrading the output SNR compared to the binaural MWF and the MWF-ITF. Furthermore, contrary to the binaural MWF and MWF-ITF, the proposed algorithm retains the spatial separation between the output speech and noise components while the binaural cues of the speech component are only slightly distorted, such that the binaural hearing advantage for speech intelligibility can still be exploited.**

***Index Terms*—Binaural cues, hearing aids, interaural coherence, multi-channel wiener filter (MWF), noise reduction.**

I. Introduction

NOISE reduction algorithms in hearing aids are crucial to improve speech understanding in background noise for hearing-impaired persons. For binaural hearing aids, algorithms that use the microphone signals from both the left and the right hearing aid are considered to be promising techniques for noise reduction, because the spatial information captured by all microphones can be exploited [1], [2]. In addition to reducing noise and limiting speech distortion, another important objective of binaural noise reduction algorithms is the preservation of the listener's impression of the acoustical scene, in order to exploit the binaural hearing advantage and to avoid confusions due to a mismatch between the acoustical and the visual information. This can be achieved by preserving the binaural cues, i.e. the Interaural Level Difference (ILD), the Interaural Time Difference (ITD) and the Interaural Coherence (IC) of all sound sources in the acoustical scene. The ILD and ITD cues are important for sound source localization [3], where the ITD cue plays a dominant role at low frequencies and the ILD cue plays a dominant role at high frequencies [4]. In addition, the IC (and its absolute value) is important for source localization in multi-source and reverberant environments since it determines the reliability of the ILD and ITD cues [5], [6]. Furthermore, the IC is an important cue for the perception of the width of sound fields [7] and is widely used in the context of room acoustics and spatial audio reproduction systems [8], [9]. Furthermore, the binaural cues ITD, ILD and IC are commonly used to predict the binaural hearing advantage, which occurs due to a spatial separation between the speech and the noise component [10], [11], in order to predict speech intelligibility for binaural signals [12], [13]. For a speech source in a low reverberant environment, masked by spherically isotropic noise, it has been shown in [14] that an improvement of the Speech Reception Threshold (SRT) for binaural hearing compared to monaural hearing up to 3.4 dB can be achieved. On the other hand, no binaural SRT increase is observed if the noise component is a interfering source located at the same position as the speech source, i.e., they contain the same spatial information [11]. Hence, for a speech source in a spherically isotropic noise field an SRT improvement up to 3.4 dB can be achieved by exploiting the binaural hearing advantage, due to the spatial separation between the speech and the noise component.

To achieve binaural noise reduction with binaural cue preservation, two main concepts have been developed. In the first concept, the microphone signals from both hearing aids are used to calculate a real-valued time-varying spectral gain, where the same gain is applied to the reference microphone signal in the left, respectively right, hearing aid [15]–[21]. This processing strategy allows for perfect preservation of the instantaneous binaural cues of both the speech and the noise component, but

Manuscript received March 11, 2015; revised July 15, 2015; accepted August 06, 2015. Date of publication August 20, 2015; date of current version September 04, 2015. This work was supported in part by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development and the Cluster of Excellence 1077 "Hearing4All," funded by the German Research Foundation(DFG). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Bozena Kostek.

The authors are with the Department of Medical Physics and Acoustics and the Cluster of Excellence "Hearing4All," University of Oldenburg, 26111 Oldenburg, Germany (e-mail: daniel.marquardt@uni-oldenburg.de; volker.hohmann@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.2015.2471096

typically suffers from a limited noise reduction performance and possible single-channel noise reduction artifacts [22]. The second concept is to apply a complex-valued filter vector to all available microphone signals on the left and the right hearing aid, combining spatial and spectral filtering [23], [24]. Using this processing strategy, a large noise reduction performance can be achieved, but the binaural cues of the noise component are not guaranteed to be preserved. Since typically the binaural cues of the speech component are preserved while the binaural cues of the residual noise component are distorted, algorithms have been proposed that aim to preserve the binaural cues of the residual noise component by including a cue preservation term in the binaural noise reduction cost function [23]–[28].

In [24] the binaural Multi-channel Wiener Filter (MWF) has been presented yielding a minimum mean square error (MMSE) estimate of the speech component in a reference microphone signal at the left and the right hearing aid. It has been theoretically proven in [23] that in the case of a single speech source the binaural MWF preserves the so-called Interaural Transfer Function (ITF), comprising the ILD and ITD cues, of the speech component, but typically distorts the ITF of the noise component, since both output components exhibit the ITF of the speech component. Hence, after applying the binaural MWF no spatial separation between the output speech and noise components exists, such that both components are perceived as coming from the same direction and no binaural unmasking can be exploited by the auditory system. In [23] an extension of the MWF, namely the MWF-ITF, has been proposed by adding a term related to the preservation of the ITF of the noise component to the binaural MWF cost function. It has been shown that a better preservation of the ITF of the noise component can be achieved, depending on the output SNR and a trade-off parameter. Instead of using a soft ITF constraint as in [23], it has been proposed in [27] to impose a hard (instantaneous) ITF constraint, however requiring an accurate estimate of the noise component. In addition in [26] an interference rejection constraint for the binaural linearly constrained minimum variance (BLCMV) beamformer has been proposed. It should however be noted that all mentioned binaural MWF extensions, aiming to preserve the binaural cues of the noise component, are only suited for *directional* noise sources since the spatial characteristics of directional sources can be well described by the ITF.

On the other hand, for *spatially isotropic* noise, whose spatial characteristics can not be properly described by the ITF but rather by the Interaural Coherence (IC), these extensions are not able to preserve the spatial characteristics. Hence another extension of the binaural MWF, namely the MWF-IC, has been proposed in [28], [29] aiming at preserving the IC of the residual noise component for spatially isotropic noise fields. Since for the MWF-IC a substantial trade-off between IC preservation and output SNR exists, in this paper we propose to control the amount of IC preservation based on the IC discrimination ability of the human auditory system. Based on psychoacoustic experimental results in [30], [31], we first define frequency-dependent upper and lower boundaries for the magnitude squared coherence (MSC) of the output noise component in order to maintain the spatial impression of the spherically isotropic noise field. Different procedures are then proposed to determine the trade-off parameter in the MWF-IC considering these boundaries, such that an optimal trade-off between spatial awareness preservation and noise reduction performance is obtained. Since the main objective of this paper is the comparison of the noise reduction and binaural cue preservation performance of MWF-based algorithms in diffuse noise fields, the algorithms are evaluated using so-called batch processing, where the noise correlation matrix is estimated from a noise-only period, and the speech-plus-noise correlation matrix is estimated during speech activity. For an online implementation the correlation matrices would need to be estimated adaptively (e.g. based on a voice activity detection mechanism), which will not be considered in this paper. However, it should be noted that even in batch processing estimation errors occur (cf. Section VI-A) and that an online version of the MWF-IC, exploiting short-term estimates of the signal statistics, has already been successfully applied in the context of dereverberation [32].

The paper is structured as follows. In Section II the configuration and notation of the used binaural hearing aid setup is described. In Section III two binaural noise reduction algorithms, namely the MWF and the MWF-ITF, are briefly reviewed and their theoretical performance in preserving the ITF and the IC of the speech and the noise component is investigated. In Section IV, the theory of spatially isotropic noise fields is reviewed and the capabilities of the MWF and the MWF-ITF in preserving the binaural cues of spatially isotropic noise fields is investigated. In Section V, the MWF-IC is presented and several procedures for determining the parameter, trading off noise reduction and IC preservation, based on psychoacoustically determined MSC boundaries are proposed. In Section VI the performance of the proposed MWF-IC and the MWF and the MWF-ITF are evaluated in terms of intelligibility-weighted SNR improvement, MSC error and ILD and ITD distributions. Experimental results in different diffuse noise scenarios show that incorporating the psychoacoustically determined MSC boundaries into the determination of the trade-off parameter for the MWF-IC yields a controllable IC preservation without significantly degrading the output SNR compared to the binaural MWF and the MWF-ITF, while retaining the spatial separation between the output speech and noise components.

II. Configuration and Notation

A. Microphone Signals and Output Signals

We consider the binaural hearing aid configuration depicted in Fig. 1, where both hearing aids are equipped with a microphone array consisting of M microphones. The m-th microphone signal in the left hearing aid $Y_{0,m}(\omega)$ can be written in the frequency-domain as

$$Y_{0,m}(\omega) = X_{0,m}(\omega) + V_{0,m}(\omega), \quad m = 1 \dots M, \qquad (1)$$

where $X_{0,m}(\omega)$ represents the speech component and $V_{0,m}(\omega)$ represents the noise component. Similarly, the m-th microphone signal in the right hearing aid can be written as $Y_{1,m}(\omega) = X_{1,m}(\omega) + V_{1,m}(\omega)$. For conciseness, we will omit the frequency-domain variable ω, except when explicitly required.

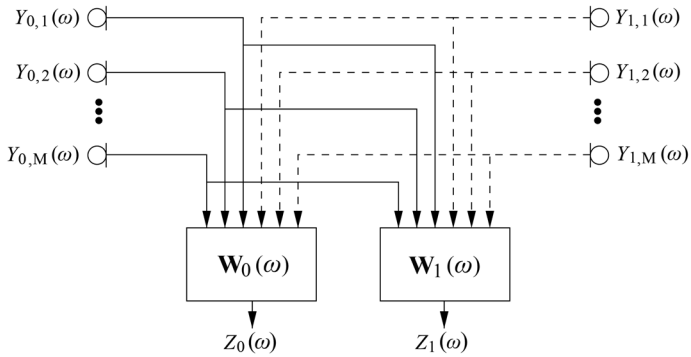


Fig. 1. General binaural processing scheme.

We define the M-dimensional stacked vectors $\mathbf{Y}_0$ and $\mathbf{Y}_1$ and the $2M$-dimensional signal vector $\mathbf{Y}$ as

$$\mathbf{Y}_0 = \begin{bmatrix} Y_{0,1} \\ \vdots \\ Y_{0,M} \end{bmatrix}, \quad \mathbf{Y}_1 = \begin{bmatrix} Y_{1,1} \\ \vdots \\ Y_{1,M} \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_1 \end{bmatrix}. \quad (2)$$

The signal vector can be written as $\mathbf{Y} = \mathbf{X}+\mathbf{V}$, where $\mathbf{X}$ and $\mathbf{V}$ are defined similarly as $\mathbf{Y}$ in (2). The correlation matrix $\mathbf{R}_y$, the speech correlation matrix $\mathbf{R}_x$ and the noise correlation matrix $\mathbf{R}_v$ are defined as

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\}, \quad \mathbf{R}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\}, \quad \mathbf{R}_v = \mathcal{E}\{\mathbf{V}\mathbf{V}^H\}, \quad (3)$$

where $\mathcal{E}$ denotes the expected value operator. Assuming that the speech and the noise components are uncorrelated, $\mathbf{R}_y$ can be written as $\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_v$.

In the case of a single speech source, the speech signal vector can be written as

$$\mathbf{X} = \mathbf{A}S, \quad (4)$$

where the $2M$-dimensional steering vector $\mathbf{A}$ contains the acoustic transfer functions from the speech source to the microphones, capturing the microphone characteristics, the head shadow effect and reverberation, and S denotes the speech signal. The speech correlation matrix then is a rank-1 matrix, i.e.

$$\mathbf{R}_x = P_s\mathbf{A}\mathbf{A}^H, \quad (5)$$

with $P_s = \mathcal{E}\{|S|^2\}$ the power spectral density of the speech signal.

Without loss of generality, we will use the first microphone on the left hearing aid and the first microphone on the right hearing aid as the so-called reference microphones for the considered speech enhancement algorithms. For conciseness, the reference microphone signals $Y_{0,1}$ and $Y_{1,1}$ of the left and the right hearing aid are denoted as Y_0 and Y_1, i.e.,

$$Y_0 = \mathbf{e}_0^T\mathbf{Y}, \quad Y_1 = \mathbf{e}_1^T\mathbf{Y}, \quad (6)$$

where $\mathbf{e}_0$ and $\mathbf{e}_1$ are $2M$-dimensional vectors with one element equal to 1 and the other elements equal to 0, i.e. $\mathbf{e}_0(1) = 1$ and $\mathbf{e}_1(M+1) = 1$. The reference microphone signals can be written as $Y_0 = X_0 + V_0$ and $Y_1 = X_1 + V_1$. The output signals Z_0 and Z_1 at the left and the right hearing aid are obtained by filtering and summing all microphone signals from both hearing aids, i.e.

$$Z_0 = \mathbf{W}_0^H\mathbf{Y}, \quad Z_1 = \mathbf{W}_1^H\mathbf{Y}, \quad (7)$$

where $\mathbf{W}_0$ and $\mathbf{W}_1$ are $2M$-dimensional complex-valued vectors. The output signal at the left hearing aid can be written as

$$Z_0 = Z_{x0} + Z_{v0} = \mathbf{W}_0^H\mathbf{X} + \mathbf{W}_0^H\mathbf{V}, \quad (8)$$

where Z_{x0} represents the speech component and Z_{v0} represents the noise component of the output signal. The output signal at the right hearing aid can be written similarly, i.e. $Z_1 = Z_{x1} + Z_{v1}$. The $4M$-dimensional complex-valued stacked vector $\mathbf{W}$ is defined as

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_0 \\ \mathbf{W}_1 \end{bmatrix}. \quad (9)$$

B. Performance Measures

In this section, we define a number of instrumental performance measures that will be used throughout the paper. The narrowband *input SNR* at the left and the right hearing aid is defined as the power ratio of the speech and noise components of the reference microphone signals, i.e.

$$SNR_0^{\text{in}} = \frac{\mathcal{E}\{|X_0|^2\}}{\mathcal{E}\{|V_0|^2\}} = \frac{\mathbf{e}_0^T\mathbf{R}_x\mathbf{e}_0}{\mathbf{e}_0^T\mathbf{R}_v\mathbf{e}_0}, \quad (10)$$

$$SNR_1^{\text{in}} = \frac{\mathcal{E}\{|X_1|^2\}}{\mathcal{E}\{|V_1|^2\}} = \frac{\mathbf{e}_1^T\mathbf{R}_x\mathbf{e}_1}{\mathbf{e}_1^T\mathbf{R}_v\mathbf{e}_1}. \quad (11)$$

The narrowband *output SNR* at the left and the right hearing aid is defined as the power ratio of the output speech and noise components, i.e.

$$SNR_0^{\text{out}} = \frac{\mathcal{E}\{|Z_{x0}|^2\}}{\mathcal{E}\{|Z_{v0}|^2\}} = \frac{\mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_0}{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0}, \quad (12)$$

$$SNR_1^{\text{out}} = \frac{\mathcal{E}\{|Z_{x1}|^2\}}{\mathcal{E}\{|Z_{v1}|^2\}} = \frac{\mathbf{W}_1^H\mathbf{R}_x\mathbf{W}_1}{\mathbf{W}_1^H\mathbf{R}_v\mathbf{W}_1}. \quad (13)$$

The *input interaural transfer function (ITF)* of the speech component and the noise component is defined as [23]

$$ITF_x^{\text{in}} = \frac{\mathcal{E}\{X_0X_1^*\}}{\mathcal{E}\{|X_1|^2\}} = \frac{\mathbf{e}_0^T\mathbf{R}_x\mathbf{e}_1}{\mathbf{e}_1^T\mathbf{R}_x\mathbf{e}_1}, \quad (14)$$

$$ITF_v^{\text{in}} = \frac{\mathcal{E}\{V_0V_1^*\}}{\mathcal{E}\{|V_1|^2\}} = \frac{\mathbf{e}_0^T\mathbf{R}_v\mathbf{e}_1}{\mathbf{e}_1^T\mathbf{R}_v\mathbf{e}_1}. \quad (15)$$

In the case of a single speech source, the input ITF in (14) is equal to the *relative transfer function (RTF)* of the speech source between the reference microphones, i.e.

$$ITF_x^{\text{in}} = RTF_x^{\text{in}} = \frac{A_0}{A_1}. \quad (16)$$

The *input interaural level difference (ILD)* of the speech component and the noise component is defined as the power ratios of the components in the left and the right hearing aid [23], i.e.

$$ILD_x^{\text{in}} = \frac{\mathcal{E}\{|X_0|^2\}}{\mathcal{E}\{|X_1|^2\}} = \frac{\mathbf{e}_0^T\mathbf{R}_x\mathbf{e}_0}{\mathbf{e}_1^T\mathbf{R}_x\mathbf{e}_1}, \quad (17)$$

$$ILD_v^{\text{in}} = \frac{\mathcal{E}\{|V_0|^2\}}{\mathcal{E}\{|V_1|^2\}} = \frac{\mathbf{e}_0^T\mathbf{R}_v\mathbf{e}_0}{\mathbf{e}_1^T\mathbf{R}_v\mathbf{e}_1}. \tag{18}$$

In the case of a single speech source, the input ILD is equal to the squared absolute value of the RTF, i.e.

$$ILD_x^{\text{in}} = \left|RTF_x^{\text{in}}\right|^2 = \left|\frac{A_0}{A_1}\right|^2. \tag{19}$$

The *output ITF* of the speech component and the noise component is defined as

$$ITF_x^{\text{out}} = \frac{\mathcal{E}\{Z_{x0}Z_{x1}^*\}}{\mathcal{E}\{|Z_{x1}|^2\}} = \frac{\mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_1}{\mathbf{W}_1^H\mathbf{R}_x\mathbf{W}_1}, \tag{20}$$

$$ITF_v^{\text{out}} = \frac{\mathcal{E}\{Z_{v0}Z_{v1}^*\}}{\mathcal{E}\{|Z_{v1}|^2\}} = \frac{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1}{\mathbf{W}_1^T\mathbf{R}_v\mathbf{W}_1}. \tag{21}$$

The *output ILD* of the speech component and the noise component is defined as

$$ILD_x^{\text{out}} = \frac{\mathcal{E}\{|Z_{x0}|^2\}}{\mathcal{E}\{|Z_{x1}|^2\}} = \frac{\mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_0}{\mathbf{W}_1^H\mathbf{R}_x\mathbf{W}_1}, \tag{22}$$

$$ILD_v^{\text{out}} = \frac{\mathcal{E}\{|Z_{v0}|^2\}}{\mathcal{E}\{|Z_{v1}|^2\}} = \frac{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0}{\mathbf{W}_1^T\mathbf{R}_v\mathbf{W}_1}. \tag{23}$$

The *interaural phase difference* (IPD) and *interaural time difference* (ITD) cues can be calculated from the ITF as [23]

$$IPD = \angle ITF, \tag{24}$$

$$ITD = \frac{IPD}{\omega}, \tag{25}$$

with $\angle$ denoting the phase.

The *input interaural coherence (IC)* of the speech component and the noise component is defined as the normalized cross-correlation between the reference microphone signals, i.e.

$$IC_x^{\text{in}} = \frac{\mathcal{E}\{X_0X_1^*\}}{\sqrt{\mathcal{E}\{|X_0|^2\}\mathcal{E}\{|X_1|^2\}}} = \frac{\mathbf{e}_0^T\mathbf{R}_x\mathbf{e}_1}{\sqrt{(\mathbf{e}_0^T\mathbf{R}_x\mathbf{e}_0)(\mathbf{e}_1^T\mathbf{R}_x\mathbf{e}_1)}}, \tag{26}$$

$$IC_v^{\text{in}} = \frac{\mathcal{E}\{V_0V_1^*\}}{\sqrt{\mathcal{E}\{|V_0|^2\}\mathcal{E}\{|V_1|^2\}}} = \frac{\mathbf{e}_0^T\mathbf{R}_v\mathbf{e}_1}{\sqrt{(\mathbf{e}_0^T\mathbf{R}_v\mathbf{e}_0)(\mathbf{e}_1^T\mathbf{R}_v\mathbf{e}_1)}}. \tag{27}$$

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

$$IC_x^{\text{out}} = \frac{\mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_0)(\mathbf{W}_1^H\mathbf{R}_x\mathbf{W}_1)}}, \tag{28}$$

$$IC_v^{\text{out}} = \frac{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_1}{\sqrt{(\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0)(\mathbf{W}_1^H\mathbf{R}_v\mathbf{W}_1)}}. \tag{29}$$

The *magnitude squared coherence (MSC)* is defined as the square of the absolute value of the IC, i.e.

$$MSC = |IC|^2. \tag{30}$$

In the case of a single speech source, the input IC in (26) is equal to the normalized input ITF in (16), i.e.,

$$IC_x^{\text{in}} = \frac{ITF_x^{\text{in}}}{|ITF_x^{\text{in}}|} = e^{j\angle ITF_x^{\text{in}}}, \tag{31}$$

which implies that the MSC is equal to 1.

III. Binaural MWF-based Noise Reduction Techniques

In this section we briefly review the cost functions for the binaural MWF [24] and the MWF-ITF [23] and investigate the properties of both algorithms in preserving the ITF and IC of the speech and the noise component for the special case of a single speech source.

A. Binaural Multi-Channel Wiener Filter (MWF)

The binaural MWF produces a minimum-mean-square-error (MMSE) estimate of the speech component in the reference microphone signals of both hearing aids, hence simultaneously reducing noise and limiting speech distortion [23], [24]. The binaural speech-distortion-weighted MWF cost function, estimating the speech components X_0 and X_1 in the left and the right hearing aid, is defined as

$$J_{\text{MWF}}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_0 - \mathbf{W}_0^H\mathbf{X} \\ X_1 - \mathbf{W}_1^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix}\mathbf{W}_0^H\mathbf{V} \\ \mathbf{W}_1^H\mathbf{V}\end{bmatrix}\right\|^2\right\}, \tag{32}$$

where the parameter μ enables a trade-off between noise reduction and speech distortion. The filter minimizing $J_{\text{MWF}}(\mathbf{W})$ is equal to [23], [24]

$$\mathbf{W}_{\text{MWF}} = \mathbf{R}^{-1}\mathbf{r}_x, \tag{33}$$

with

$$\mathbf{R} = \begin{bmatrix}\mathbf{R}_x + \mu\mathbf{R}_v & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_x + \mu\mathbf{R}_v\end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix}\mathbf{R}_x\mathbf{e}_0 \\ \mathbf{R}_x\mathbf{e}_1\end{bmatrix}. \tag{34}$$

The filters for the left and the right hearing aid can then be written as

$$\mathbf{W}_{\text{MWF},0} = (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}\mathbf{R}_x\mathbf{e}_0, \tag{35}$$

$$\mathbf{W}_{\text{MWF},1} = (\mathbf{R}_x + \mu\mathbf{R}_v)^{-1}\mathbf{R}_x\mathbf{e}_1. \tag{36}$$

In the case of a single speech source, it has been shown in [23], [24] that, using (5), these filters can be written as

$$\boxed{\begin{aligned}\mathbf{W}_{\text{MWF},0} &= \frac{P_s}{\mu+\rho}A_0^*\mathbf{R}_v^{-1}\mathbf{A} \\ \mathbf{W}_{\text{MWF},1} &= \frac{P_s}{\mu+\rho}A_1^*\mathbf{R}_v^{-1}\mathbf{A}\end{aligned}} \tag{37}$$

with

$$\rho = P_s\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}. \tag{38}$$

Substituting (37) in (12) and (13), it can be shown that the output SNR at the left and the right hearing aid are the same and equal to

$$SNR_0^{\text{out}} = SNR_1^{\text{out}} = \rho. \tag{39}$$

Since the filters in (37) are related as

$$\mathbf{W}_{\text{MWF},0} = ITF_x^{\text{in},*}\mathbf{W}_{\text{MWF},1}, \tag{40}$$

it can be shown by substituting (40) in (20) and (21), that the output ITF of the speech and the noise component are the same and equal to ITF_x^{in}, i.e.,

$$ITF_x^{\text{out}} = ITF_v^{\text{out}} = ITF_x^{\text{in}}. \tag{41}$$

In addition, by substituting (40) in (28) and (29), we can now show that the output IC of the speech and the noise component are also the same and equal to

$$IC_x^{out} = IC_v^{\text{out}} = IC_x^{\text{in}} = e^{j\angle ITF_x^{\text{in}}}, \tag{42}$$

such that

$$MSC_x^{\text{out}} = MSC_v^{\text{out}} = 1. \tag{43}$$

From (41) and (42) we can conclude that the output ITF and IC of the speech and the noise component are the same and are equal to the input ITF and IC of the speech component, implying that both output components are perceived as directional sources coming from the speech direction, which is not desired for the preservation of the binaural cues of the noise component.

B. Binaural MWF with ITF Preservation (MWF-ITF)

In order to control the binaural cues of the noise component, it has been proposed in [23], [25] to extend the binaural MWF cost function with a term related to the ITF of the noise component. The ITF cost function for preserving the binaural cues of the noise component is defined as

$$J_{\text{ITF}}(\mathbf{W}) = \mathcal{E}\{|\mathbf{W}_0^H\mathbf{V} - ITF_v^{\text{in}}\mathbf{W}_1^H\mathbf{V}|^2\} = \mathbf{W}^H\mathbf{R}_{vt}\mathbf{W}, \tag{44}$$

with

$$\mathbf{R}_{vt} = \begin{bmatrix} \mathbf{R}_v & -ITF_v^{\text{in},*}\mathbf{R}_v \\ -ITF_v^{\text{in}}\mathbf{R}_v & |ITF_v^{\text{in}}|^2\mathbf{R}_v \end{bmatrix}. \tag{45}$$

The total cost function, trading off noise reduction, speech distortion, and binaural cue preservation, is defined as

$$J_{\text{MWF-ITF}}(\mathbf{W}) = J_{\text{MWF}}(\mathbf{W}) + \delta J_{\text{ITF}}(\mathbf{W}), \tag{46}$$

where the parameter δ enables to put more emphasis on binaural cue preservation for the noise component. The filter minimizing $J_{\text{MWF-ITF}}(\mathbf{W})$ is equal to

$$\mathbf{W}_{\text{MWF-ITF}} = (\mathbf{R} + \delta\mathbf{R}_{vt})^{-1}\mathbf{r}_x. \tag{47}$$

In the case of a single speech source, it has been shown in [23] that the stacked filter in (47) can be written as

$$\boxed{\begin{aligned} \mathbf{W}_{\text{MWF-ITF},0} &= \frac{P_s}{\mu+\rho}\left(A_0^* - \xi\right)\mathbf{R}_v^{-1}\mathbf{A} \\ \mathbf{W}_{\text{MWF-ITF},1} &= \frac{P_s}{\mu+\rho}\left(A_1^* + \xi ITF_v^{\text{in}}\right)\mathbf{R}_v^{-1}\mathbf{A} \end{aligned}} \tag{48}$$

with

$$\xi = \frac{\delta\left(A_0^* - ITF_v^{\text{in},*}A_1^*\right)}{\mu + \rho + \delta(1 + |ITF_v^{\text{in}}|^2)}. \tag{49}$$

Note that the filters in (48) are equal to the binaural MWF filters in (37) and an extra term due to the extension with the ITF cost function for the noise component. It has been shown in [23] that the narrowband output SNR for the MWF-ITF is the same as for the MWF, i.e.,

$$SNR_0^{\text{out}} = SNR_1^{\text{out}} = \rho. \tag{50}$$

Moreover, since it can be shown that the filters in (48) are related as [23]

$$\mathbf{W}_{\text{MWF-ITF},0} = ITF^{\text{out},*}\mathbf{W}_{MWF-ITF,1}, \tag{51}$$

with

$$ITF^{\text{out}} = \frac{A_0 - \xi^*}{A_1 + \xi^* ITF_v^{\text{in},*}}, \tag{52}$$

the output ITF of the speech and the noise components are again the same and equal to

$$ITF_x^{\text{out}} = ITF_v^{\text{out}} = ITF^{\text{out}}. \tag{53}$$

In addition, by substituting (51) in (28) and (29) we can now show that the output IC of the speech and the noise components are also the same and equal to

$$IC_x^{\text{out}} = IC_v^{\text{out}} = \frac{ITF^{\text{out}}}{|ITF^{\text{out}}|} = e^{j\angle ITF^{\text{out}}}, \tag{54}$$

such that

$$MSC_x^{\text{out}} = MSC_v^{\text{out}} = 1. \tag{55}$$

Similarly as for the binaural MWF, both output components for the MWF-ITF are perceived as directional sources coming from the same direction. This direction is determined by ITF^{out} in (52) and depends, e.g., on the trade-off parameter δ and the output SNR ρ. If $\delta = 0$, then $ITF^{\text{out}} = ITF_x^{\text{in}}$, and if $\delta \to \infty$, then $ITF^{\text{out}} = ITF_v^{\text{in}}$, such that there is always a trade-off between preserving the ITF of the speech component and preserving the ITF of the noise component. Using the MWF-ITF, the ITF and the IC of a directional noise source can hence be preserved at the cost of distorting the ITF and the IC of the speech component. However, as has been noted in [23], equation (52) implies that for high output SNRs the output ITF is shifted towards the ITF of the speech component and for low output SNRs the output ITF is shifted towards the ITF of the noise component. Due to this advantageous perceptual effect, an increase in localization performance for the MWF-ITF compared to the MWF has been observed in subjective listening experiments [25].

IV. Performance of the MWF and MWF-ITF in Isotropic Noise Fields

Since the objective of this paper is binaural noise reduction and cue preservation for spatially isotropic noise fields, more specifically spherically isotropic noise fields, in Section IV-A we briefly review spherically isotropic noise fields in free-field and binaural conditions. In Section IV-B we investigate the binaural cue preservation performance of the MWF and the MWF-ITF in spherically isotropic noise fields based on the theoretical analysis in Section III.

A. Spherically Isotropic Noise Fields

Spatially isotropic noise fields have been shown to be reasonable approximations for noise fields in crowded rooms. A spherically isotropic sound field is defined as a sound field that is composed of a superposition of uncorrelated plane waves that are uniformly distributed on a sphere with equal power spectrum densities [33]. The spatial coherence of a spherically isotropic noise field can be derived by integrating over all plane waves under free-field condition. The spatial coherence for 2 microphones at a distance d is equal to

$$\Gamma(\omega) = \operatorname{sinc}\left(\frac{\omega d}{c}\right), \tag{56}$$

with $\operatorname{sinc}(x) = \frac{\sin(x)}{x}$ and c denoting the speed of sound. For a binaural setup, the free-field condition is not valid, since the head shadow effect needs to be taken into account. In [34] a method was presented for calculating the interaural coherence based on a geometrical model of the human head. Based on experimental data in [35] it was shown that the IC of a spherically isotropic noise field can be approximated as a modified sinc-function, i.e.,

$$IC(\omega) = \operatorname{sinc}\left(\alpha\frac{\omega d}{c}\right)\frac{1}{\sqrt{\left(1+\left(\beta\frac{\omega d}{c}\right)^4\right)}}, \tag{57}$$

with $\alpha = 2.2$ and $\beta = 0.5$. This formula implies that the presence of the head results in a shifting of the zero crossings of the sinc-function towards lower frequencies and causes an additional frequency-dependent damping. These effects also occur in the physical model in [34]. Note that in the remainder of this paper we will refer to spherically isotropic noise fields as diffuse noise fields.

B. Performance of the MWF and the MWF-ITF

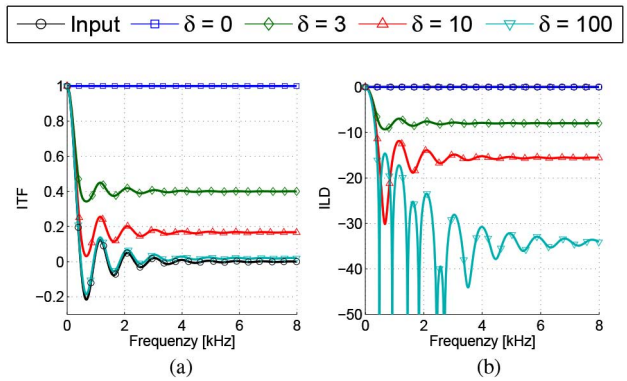
As shown in Section III, both for the MWF and for the MWF-ITF the output IC of the speech and the noise components are equal to each other, cf. (42) and (54). Moreover, the output MSC of the speech and the noise components are both equal to 1, cf. (43) and (55). Hence, also in the case of a diffuse noise field, both output speech and noise components will be perceived as directional sources from the same direction (determined by ITF^{out}), such that the perceived width of a diffuse noise field will not be present in the output noise component of the MWF and the MWF-ITF and no binaural hearing advantage can be exploited.

For a diffuse noise field the power spectral densities of the input noise components of the reference microphone signals are the same, i.e., $\mathcal{E}\{|V_0|^2\} = \mathcal{E}\{|V_1|^2\}$, such that

$$ITF_v^{in} = IC_v^{in}. \tag{58}$$

Furthermore, assuming the common hearing aid scenario of a speech source located in front of the listener and assuming symmetry of the head, i.e., $A_0 = A_1$, the output ITF of the MWF-ITF in (52) can then be computed as

$$ITF^{\text{out}} = \frac{1-\frac{\delta\left(1-IC_v^{\text{in}}\right)}{\mu+\rho+\delta(1+|IC_v^{\text{in}}|^2)}}{1+\frac{\delta IC_v^{\text{in}}(1-IC_v^{\text{in}})}{\mu+\rho+\delta(1+|IC_v^{\text{in}}|^2)}}, \tag{59}$$



(a) (b)

Fig. 2. Binaural Cues of the MWF-ITF according to (59). The parameters μ and ρ are equal to 1, $c = 340\frac{\text{m}}{\text{s}}$ and $d = 0.164$ m (a) ITF (b) ILD in dB.

where IC_v^{in} can be calculated using the IC model in (57). Substituting (48) in (23) and exploiting (51), the output ILD of the noise component can be calculated as

$$ILD_v^{\text{out}} = \left|ITF^{\text{out}}\right|^2. \tag{60}$$

Fig. 2 depicts the output ITF and ILD calculated according to (59) and (60) for several trade-off parameters δ. As expected from the theoretical analysis, the output ITF in Fig. 2(a) is real-valued and converges towards the input ITF when the trade-off parameter δ is increased. As depicted in Fig. 2(b), the output ILD (in dB) is always negative and frequency-dependent and hence significantly different from the frequency-independent input ILD which is equal to 0 dB. Due to the relation between the output IC and the output ITF in (54), the output IC is equal to 1 or -1, depending on the sign of the output ITF, i.e.,

$$IC_v^{\text{out}} = \begin{cases} 1, & \text{if} \quad ITF^{\text{out}} \geq 0 \\ -1, & \text{if} \quad ITF^{\text{out}} < 0 \end{cases}. \tag{61}$$

Consequently, the output MSC is always equal to 1. Since the IPD can be calculated from the ITF, cf. (24), the IPD is equal to 0 or π, depending on the sign of the output ITF, i.e.,

$$IPD_v^{\text{out}} = \begin{cases} 0, & \text{if} \quad ITF^{\text{out}} \geq 0 \\ \pi, & \text{if} \quad ITF^{\text{out}} < 0 \end{cases}. \tag{62}$$

Note that the same analysis also holds for the output binaural cues of the speech component due to the relation in (53). From this analysis we can conclude that in diffuse noise fields the fairly unnatural combinations of the output ILD and ITD for the MWF-ITF will lead to perceptually unsatisfying results for both the speech and the noise component, especially for large trade-off parameters δ. Due to the perceptual disadvantages of the MWF and the MWF-ITF in diffuse noise fields, in the following section we propose an extension of the binaural MWF cost function, aiming to preserve the interaural coherence.

V. MWF with Perceptually Constrained IC Preservation

Since for diffuse noise the MWF and the MWF-ITF are not able to preserve the binaural cues (i.e. the interaural coherence) of the noise component, in [28] an extension of the binaural MWF, namely the MWF-IC, has been proposed, which aims at preserving the IC of the noise component. The MWF-IC is reviewed in Section V-A. Since for the MWF-IC a

trade-off between noise reduction and IC preservation exists, in Section V-B different methods for determining the trade-off parameter based on psychoacoustic experiments are proposed, which are discussed in Section V-C.

A. MWF with IC Preservation (MWF-IC)

In [28] an extension of the binaural MWF has been proposed, by adding an IC preservation term for the noise component to the binaural MWF cost function in (32), defined as

$$J_{\mathrm{IC}}(\mathbf{W}) = \left|IC_v^{\mathrm{out}} - IC_v^{\mathrm{des}}\right|^2, \tag{63}$$

$$= \left| \frac{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_1}{\sqrt{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}} - IC_v^{\mathrm{des}} \right|^2, \tag{64}$$

where IC_v^{des} denotes the desired output IC of the noise component. The desired output IC can, e.g., be equal to the input IC of the noise component in (27) or can be defined based on models of the IC in diffuse noise fields, as discussed in Section IV-A. Since it may be difficult in practice to estimate the input IC of the noise component during speech-and-noise periods, in this paper we will use the modified sinc-function in (57) as a reasonable and pragmatic choice. Similarly as for the MWF-ITF, the total cost function, trading off noise reduction, speech distortion and IC preservation, is defined as

$$J_{\mathrm{MWF-IC}}(\mathbf{W}) = J_{\mathrm{MWF}}(\mathbf{W}) + \lambda J_{\mathrm{IC}}(\mathbf{W}), \tag{65}$$

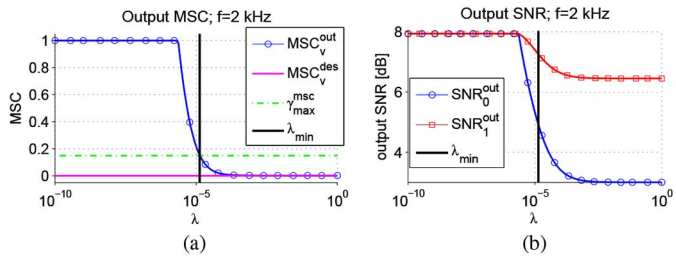
where the parameter λ is used to put more emphasis on IC preservation for the noise component. Since no closed-form expression is available for the MWF-IC filter $\mathbf{W}(\lambda)$ minimizing the non-linear cost function in (65), an iterative numerical optimization method is required, for which we have used a large-scale trust region method [36], [37]. In order to improve the numerical robustness and the convergence speed, analytical expressions for the gradient and the Hessian of the cost function $J_{\mathrm{MWF-IC}}(\mathbf{W})$ have been provided. The analytical expression for the gradient can be found in Appendix A, whereas the analytical expression for the Hessian is omitted due to space constraints.

B. Optimization of the Trade-Off Parameter λ

The parameter λ in (65) enables to trade off noise reduction and IC preservation of the noise component. Hence, it is crucial to determine a suitable (frequency-dependent) parameter λ which provides an acceptable trade-off between noise reduction and IC preservation. In order to control the amount of IC preservation, we first determine the set of possible trade-off parameters Λ, for which the MSC of the output noise component for the resulting filter $\mathbf{W}(\lambda)$ lies between the frequency-dependent boundaries $\gamma_{\mathrm{min}}^{\mathrm{msc}}$ and $\gamma_{\mathrm{max}}^{\mathrm{msc}}$, i.e.,

$$\Lambda = \left\{\lambda | \gamma_{\mathrm{min}}^{\mathrm{msc}} \leq MSC_v^{\mathrm{out}}(\mathbf{W}(\lambda)) \leq \gamma_{\mathrm{max}}^{\mathrm{msc}}\right\}, \tag{66}$$

with $MSC_v^{\mathrm{out}} = |IC_v^{\mathrm{out}}|^2$ and IC_v^{out} defined in (29). We propose to impose a constraint on the MSC, since psychoacoustical experiments have shown that the perceived width of a sound source mainly depends on the absolute value of the IC [7]. The choice of the boundaries $\gamma_{\mathrm{min}}^{\mathrm{msc}}$ and $\gamma_{\mathrm{max}}^{\mathrm{msc}}$ based on psychoacoustically motivated criteria will be discussed in detail



(a) (b)

Fig. 3. Output MSC of the noise component and output SNR for the left and the right hearing aid for different values of the trade-off parameter λ for a frequency of 2 kHz. The vertical line indicates the minimal value λ_{min} for which the MSC constraints $\gamma_{\mathrm{max}}^{\mathrm{msc}} = 0.16$ and $\gamma_{\mathrm{min}}^{\mathrm{msc}} = 0$ are satisfied. MSC_v^{des} is equal to 0.01 (a) Output MSC (b) Output SNR.

in Section V-C. For an exemplary scenario (cf. simulation setup in Section VI-A), Fig. 3 depicts the output MSC of the noise component and the output SNR for the left and the right hearing aid for different values of the trade-off parameter λ for a frequency of 2 kHz. As can be observed from this figure, for larger values of λ the difference between the output MSC and the desired MSC becomes smaller at the expense of a degraded output SNR. The vertical line indicates the smallest trade-off parameter λ_{min} that satisfies the MSC constraint for a certain value $\gamma_{\mathrm{max}}^{\mathrm{msc}}$ which should be chosen to be larger than MSC_v^{des}. As can be observed, all parameters $\lambda > \lambda_{\mathrm{min}}$ also satisfy the MSC constraint but result in lower output SNRs, whereas all parameters $\lambda < \lambda_{\mathrm{min}}$ result in larger output SNRs but do not satisfy the MSC constraint.

Assuming monotonically decreasing output SNRs as depicted in Fig. 3, which is typically - but not always - the case, the smallest value of the parameter λ satisfying the MSC constraint in (66), i.e.,

$$\lambda_{\mathrm{min}} = \min_{\lambda \in \Lambda}(\lambda), \tag{67}$$

will result in the largest output SNR for both hearing aids.

However, since it can not be theoretically proven that the output SNRs are monotonically decreasing, we also propose two other procedures for determining the trade-off parameter, the first one optimizing the narrowband average output SNR, i.e.,

$$\lambda_{\mathrm{snr}} = \arg\max_{\lambda \in \Lambda} \frac{SNR_0^{\mathrm{out}}(\mathbf{W}(\lambda)) + SNR_1^{\mathrm{out}}(\mathbf{W}(\lambda))}{2}, \tag{68}$$

and the second one optimizing the narrowband better ear output SNR, i.e.,

$$\lambda_{\mathrm{snr,be}} = \arg\max_{\lambda \in \Lambda} SNR_{\mathrm{be}}^{\mathrm{out}}(\mathbf{W}(\lambda)), \tag{69}$$

with

$$SNR_{\mathrm{be}}^{\mathrm{out}}(\mathbf{W}(\lambda)) = \max\left[SNR_0^{\mathrm{out}}(\mathbf{W}(\lambda)), SNR_1^{\mathrm{out}}(\mathbf{W}(\lambda))\right]. \tag{70}$$

To determine these trade-off parameters, we have used an *exhaustive search* method. Since this is a computationally expensive method, we also propose an *iterative search* method resulting in the trade-off parameter λ_{it}. The iterative search method is initialized with a large value λ_{init} such that the MSC constraint is definitely satisfied. This value is repeatedly decreased by a factor 10 until the MSC constraint in (66) is not satisfied. The trade-off parameter λ is then increased by this value

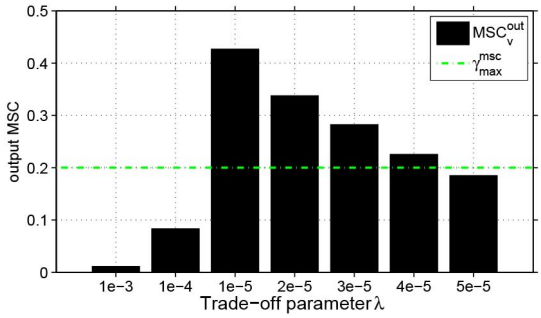


Fig. 4. Iterative search method for the trade-off parameter λ_{it}. The iterative search method is initialized with $\lambda_{init} = 1e-3$ which is decreased by a factor 10 until the MSC constraint is not satisfied ($\lambda = 1e-5$). The trade-off parameter is then increased by this value ($1e-5$) until the MSC constraint is again satisfied.

until the MSC constraint is again satisfied. The output MSC and the trade-off parameters for this iterative search method are exemplarily depicted in Fig. 4.

C. Psychoacoustically Motivated MSC Boundaries $\gamma_{\min}^{\mathrm{msc}}$ and $\gamma_{\max}^{\mathrm{msc}}$

The psychoacoustically motivated constraint boundaries $\gamma_{\min}^{\mathrm{msc}}$ and $\gamma_{\max}^{\mathrm{msc}}$ in (66) can be defined based on subjective listening experiments evaluating the IC discrimination abilities of the human auditory system in a diffuse noise field. In [30] frequency-dependent IC discrimination thresholds in a diffuse noise field have been measured for frequencies up to 1.5 kHz. It has been shown that the sensitivity to changes in the IC from a reference IC strongly depends on the reference IC value. For a reference IC close to 1 small changes can be perceived, whereas for a reference IC close to 0 the human auditory system is less sensitive to changes in the IC. This is consistent with the perceptual results of other IC discrimination studies [38], [39]. Furthermore, in [31] the IC discrimination sensitivity in a diffuse noise field was examined by setting the reference IC below 500 Hz equal to 1 and the reference IC above 500 Hz equal to 0, approximating the IC of a diffuse noise field. The perceptual results indicate that for frequencies above 500 Hz a deviation of the IC of ± 0.6 is not discriminable from the reference IC of 0.

Combining the subjective results from [30] and [31], we define the following constraint boundaries $\gamma_{\min}^{\mathrm{msc}}$ and $\gamma_{\max}^{\mathrm{msc}}$.

For frequencies below 500 Hz, the boundaries $\gamma_{\min}^{\mathrm{msc}}$ and $\gamma_{\max}^{\mathrm{msc}}$ are a function of the desired MSC, denoted as $g(MSC_v^{\mathrm{des}})$ and $h(MSC_v^{\mathrm{des}})$, respectively, according to the results in [30]. The functions $g(x)$ and $h(x)$ can be approximated from the discrete data in [30] using polynomial fitting, with

$$g(x) = \begin{cases} 2.88x^2 - 2.96x + 0.715, & x \in [0.64 \dots 1] \\ 0, & \text{else} \end{cases} \tag{71}$$

$$h(x) = 0.78x^3 - 1.76x^2 + 1.57x + 0.42, x \in [0 \dots 1] \tag{72}$$

For frequencies above 500 Hz, we define an MSC-independent lower and upper boundary inspired by the perceptual results in [31]. The lower boundary $\gamma_{\min}^{\mathrm{msc}}$ is set to 0. To investigate the impact of the MSC-independent upper boundary on the output

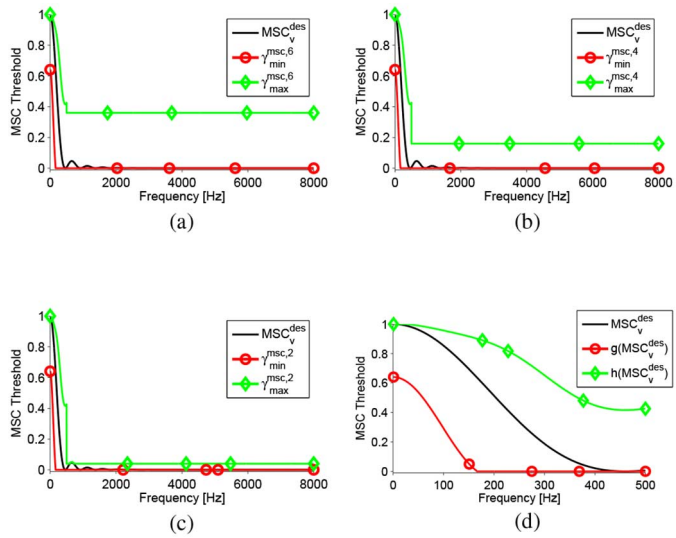


Fig. 5. Psychoacoustically motivated lower and upper MSC boundaries. For frequencies below 500 Hz, the boundaries depend on $MSC_{\mathrm{v}}^{\mathrm{des}}$. For frequencies above 500 Hz, the boundaries are independent of $MSC_{\mathrm{v}}^{\mathrm{des}}$. (a) MSC boundaries $\gamma_{\min}^{\mathrm{msc},6}$ and $\gamma_{\max}^{\mathrm{msc},6}$. (b) MSC boundaries $\gamma_{\min}^{\mathrm{msc},4}$ and $\gamma_{\max}^{\mathrm{msc},4}$. (c) MSC boundaries $\gamma_{\min}^{\mathrm{msc},2}$ and $\gamma_{\max}^{\mathrm{msc},2}$. (d) Boundary functions $g(MSC_{\mathrm{v}}^{\mathrm{des}})$ and $h(MSC_{\mathrm{v}}^{\mathrm{des}})$ for $f \leq 500$ Hz.

SNR and the IC preservation, we consider three boundary values for the MSC, i.e., $\gamma_{\max}^{\mathrm{msc},6} = 0.36$, $\gamma_{\max}^{\mathrm{msc},4} = 0.16$, $\gamma_{\max}^{\mathrm{msc},2} = 0.04$, corresponding to an IC threshold of ± 0.6, ± 0.4 and ± 0.2. The lower and upper MSC boundaries $\gamma_{\min}^{\mathrm{msc}}$ and $\gamma_{\max}^{\mathrm{msc}}$ which will be used in the following simulations are hence defined as

$$\gamma_{\min}^{\mathrm{msc},6\backslash 4\backslash 2} = \begin{cases} g(MSC_v^{\mathrm{des}}) & , f \leq 500Hz \\ 0 & , f > 500Hz \end{cases}, \tag{73}$$

$$\gamma_{\max}^{\mathrm{msc},6\backslash 4\backslash 2} = \begin{cases} h(MSC_v^{\mathrm{des}}) & , f \leq 500Hz \\ 0.36\backslash 0.16\backslash 0.04 & , f > 500Hz \end{cases}, \tag{74}$$

and are depicted in Fig. 5. To illustrate the MSC dependency of the lower and upper boundaries for frequencies below 500 Hz, the functions $g(MSC_{\mathrm{v}}^{\mathrm{des}})$ and $h(MSC_{\mathrm{v}}^{\mathrm{des}})$ for frequencies below 500 Hz are depicted in Fig. 5(d). Based on the subjective listening experiments in [30] and [31], it is hence assumed that if the output MSC lies within the MSC boundaries in Fig. 5, the spatial impression of the output noise component is perceptually not discriminable from the spatial impression of a diffuse noise field.

VI. Simulations

In this section, we present several simulation results for a real-world cafeteria scenario, comparing the performance of the proposed binaural noise reduction and cue preservation algorithms in terms of instrumental measures. In the first experiment, we compare the different procedures for selecting the trade-off parameter λ in the MWF-IC as introduced in Section V-B. These results are then used in the second experiment, where we compare the performance of the MWF, the MWF-ITF and the proposed MWF-IC using psychoacoustically motivated MSC boundaries.

A. Input Signals and Signal Statistics

The hearing aid microphone signals have been generated using measured impulse responses for a binaural hearing aid

setup mounted on a dummy head in a cafeteria with a reverberation time of about 1250 ms [40]. Each hearing aid was equipped with 2 microphones with a distance of 0.7 cm. The distance between the microphones of the left and the right hearing aid was approximately 0.164 m. The speech source was located in front of the dummy head at a distance of 1 m. Two different noise types have been used for the experiments:

- *Babble noise:* To allow for a controlled experiment, a spatially stationary noise field was generated using the method described in [41], where the time-varying power spectral density (PSD) of the noise component was calculated from a babble noise signal and the time-invariant spatial coherence matrix of the binaural setup was calculated according to (57) with $c = 340\frac{m}{s}$ and the distance between the microphones d was equal to the values stated above.
- *Ambient noise:* To generate a more realistic scenario, recorded ambient noise from the cafeteria including babble noise, clacking plates and interfering speakers has been used as the noise component [40].

The speech-and-noise signals had a length of 10 s and were preceded by a noise-only signal of 3 s. The speech and noise components were mixed such that the average intelligibility-weighted input SNR [42] in the reference microphones at the left and the right hearing aid was equal to -5, 0, and 5 dB, respectively, at a sampling frequency of 16 kHz. The microphone signals were transformed to the frequency-domain using the short-time Fourier transform (STFT) with segments of length N overlapping by $N-L$ samples, e.g., for the reference microphone signal of the left hearing aid

$$\begin{aligned} Y_0(k,i) &= \sum_{n=0}^{N-1} y_0(iL+n)w(n)e^{-j\Omega_k n}, \\ &= X_0(k,i) + V_0(k,i), \end{aligned} \tag{75}$$

with k the frequency index, i the block index, $\Omega_k = 2\pi k/N$ the normalized angular frequency, and $w(n)$ an analysis window of length N. The segment length was set to $N = 512$, L was set to 256 and for the analysis window we have used a Hann window.

For calculating the filter vectors $\mathbf{W}$ for the MWF, MWF-ITF and MWF-IC, the correlation matrices of the signal components are estimated as

$$\hat{\mathbf{R}}_y(k) = \frac{1}{L_y}\sum_{i=0}^{L_y-1} \mathbf{Y}(k,i)\mathbf{Y}^H(k,i) \quad \text{speech-and-noise}, \tag{76}$$

$$\hat{\mathbf{R}}_v(k) = \frac{1}{L_v}\sum_{i=0}^{L_v-1} \mathbf{V}(k,i)\mathbf{V}^H(k,i) \quad \text{noise-only}, \tag{77}$$

where the speech-plus-noise correlation matrix $\hat{\mathbf{R}}_y(k)$ has been computed during the 10 s speech-and-noise part and the noise correlation matrix $\hat{\mathbf{R}}_v(k)$ has been computed during the 3 s noise-only part. L_y denotes the number of segments during the 10 s speech-and-noise part and L_v denotes the number of segments during the 3 s noise-only part. Since the speech component is not available in real-world scenarios, the speech correlation matrix has been estimated as

$$\hat{\mathbf{R}}_x(k) = \hat{\mathbf{R}}_y(k) - \hat{\mathbf{R}}_v(k). \tag{78}$$

Since the noise correlation matrix is estimated during the noise-only period, it will deviate from the noise correlation matrix during the speech-and-noise period. In addition, by estimating the speech correlation matrix as in (78), additional estimation errors are introduced. Since due to this estimation errors it can not be guaranteed that the speech correlation matrix estimate $\hat{\mathbf{R}}_x(k)$ is positive semi-definite, which may lead to signal distortions, we have used the rank-1 approximation

$$\hat{\mathbf{R}}_x^1(k) = \sigma_1(k)\mathbf{q}_1(k)\mathbf{q}_1^H(k), \tag{79}$$

with $\sigma_1(k)$ the largest eigenvalue of $\hat{\mathbf{R}}_x(k)$ and $\mathbf{q}_1(k)$ the corresponding eigenvector. In the case of estimation errors, using a rank-1 approximation has been shown to improve the output SNR [29], [43]. For the MWF-ITF the input ITF of the noise component $ITF_{\mathrm{v}}^{\mathrm{in}}$ is calculated according to (15) using the estimated noise correlation matrix $\hat{\mathbf{R}}_v$. For the MWF-IC the desired IC for the noise component $IC_{\mathrm{v}}^{\mathrm{des}}$ is calculated according to (57) with $c = 340\frac{\mathrm{m}}{\mathrm{s}}$ and $d = 0.164$ m.

The narrowband output SNR at the left and the right hearing aid is calculated according to (12) and (13) and the intelligibility-weighted broadband output SNR (iSNR) [42] is calculated as

$$iSNR^{\mathrm{out}} = \sum_{k=1}^{N-1} I(k)10\log_{10}(SNR^{\mathrm{out}}(k)), \tag{80}$$

where $I(k)$ is a weighting function that takes the importance of different frequency bands for speech intelligibility into account. Since the better ear SNR mainly determines speech intelligibility, especially in the case when a large SNR difference between the left and the right ear occurs [44], we also calculate the broadband better ear iSNR as

$$iSNR_{\mathrm{be}}^{\mathrm{out}} = \max(iSNR_0^{\mathrm{out}}, iSNR_1^{\mathrm{out}}). \tag{81}$$

To evaluate the binaural cue preservation for the noise component, we have used the broadband MSC error ΔMSC_v, which has been calculated by averaging the MSC error across frequencies, i.e.,

$$\Delta MSC_v = \frac{1}{N-1}\sum_{k=1}^{N-1} \left|MSC_v^{\mathrm{des}}(k) - MSC_v^{\mathrm{out}}(k)\right|, \tag{82}$$

where the output MSC has been calculated according to (30) and the output IC has been calculated according to (29).

For the directional speech component the MSC error is however not an appropriate objective measure, since the MSC contains information about the amount of correlation of a signal in the microphones but does not contain information about the perceived direction of a directional source. Hence, to evaluate the binaural cue preservation of the speech component we calculate the distribution of the so-called reliable ILD and ITD cues using a model of binaural auditory processing [6]. For the distributions of the reliable ITD cues only frequencies up to 1.5 kHz have been used as described in [6].

TABLE I
AVERAGE NORMALIZED ESTIMATION ERROR FOR THE SPEECH AND THE NOISE CORRELATION MATRICES FOR DIFFERENT INPUT iSNRs AND NOISE TYPES

	-5 dB	0 dB	5 dB
$\Delta\mathbf{R}_x$ Babble [dB]	-10.2	-18.3	-25.2
$\Delta\mathbf{R}_v$ Babble [dB]	-10.8	-10.8	-10.8
$\Delta\mathbf{R}_x$ Ambient [dB]	-7.0	-15.3	-23.0
$\Delta\mathbf{R}_v$ Ambient [dB]	-9.4	-9.4	-9.4

In order to quantify the estimation errors of the correlation matrices, we have considered the normalized estimation error of the speech and the noise correlation matrices, i.e.,

$$\Delta\mathbf{R}_x(k) = 10\log_{10}\frac{\left\|\mathbf{R}_x(k) - \hat{\mathbf{R}}_x^1(k)\right\|_2^2}{\|\mathbf{R}_x(k)\|_2^2}, \tag{83}$$

$$\Delta\mathbf{R}_v(k) = 10\log_{10}\frac{\left\|\mathbf{R}_v(k) - \hat{\mathbf{R}}_v(k)\right\|_2^2}{\|\mathbf{R}_v(k)\|_2^2}, \tag{84}$$

where $\mathbf{R}_x(k)$ and $\mathbf{R}_v(k)$ denote the speech and the noise correlation matrices calculated during the speech-and-noise period, and $\hat{\mathbf{R}}_x^1(k)$ and $\hat{\mathbf{R}}_v(k)$ denote the estimated speech and noise correlation matrices calculated according to (79) and (77). For the input iSNRs of -5, 0 and 5 dB and the two considered noise types (babble noise and ambient noise), Table I shows the normalized estimation errors averaged over frequency. While the normalized estimation error of the noise correlation matrix is independent of the iSNR, the normalized estimation error of the speech correlation matrix decreases with increasing iSNR. Furthermore, the normalized estimation error of the speech correlation matrix is larger for the ambient noise type than for the babble noise type due to the spatial non-stationarity. It should be furthermore noted that due to the batch processing, the trade-off parameter λ is also determined based on the long-term estimates of the speech and noise correlation matrices. In an online implementation with time-varying estimates of the speech and the noise correlation matrices, in principle the same procedure as described in Section V-B could be used, leading to a time-varying trade-off parameter λ. However, due to the batch processing this has not been considered in this paper.

B. Experimental Results

1) Experiment 1: In the first experiment, we compare the different methods proposed in Section V-B for selecting the trade-off parameter λ in the MWF-IC that yields a desired output MSC for the noise component. The ambient noise was added to the speech component such that the average intelligibility-weighted input SNR in the reference microphones at the left and the right hearing aid was equal to 0 dB. For the exhaustive search methods we have used 500 values for λ, which are logarithmically spaced between 10^{-10} and 1. The iterative search method has been initialized with $\lambda_{\text{init}} = 10$. To have a realistic procedure for determining the trade-off parameter λ, in this experiment the output iSNR and MSC error have been computed based on the same correlation matrices that have been used to compute the filter vectors, i.e., $\hat{\mathbf{R}}_x^1$ and $\hat{\mathbf{R}}_v$.

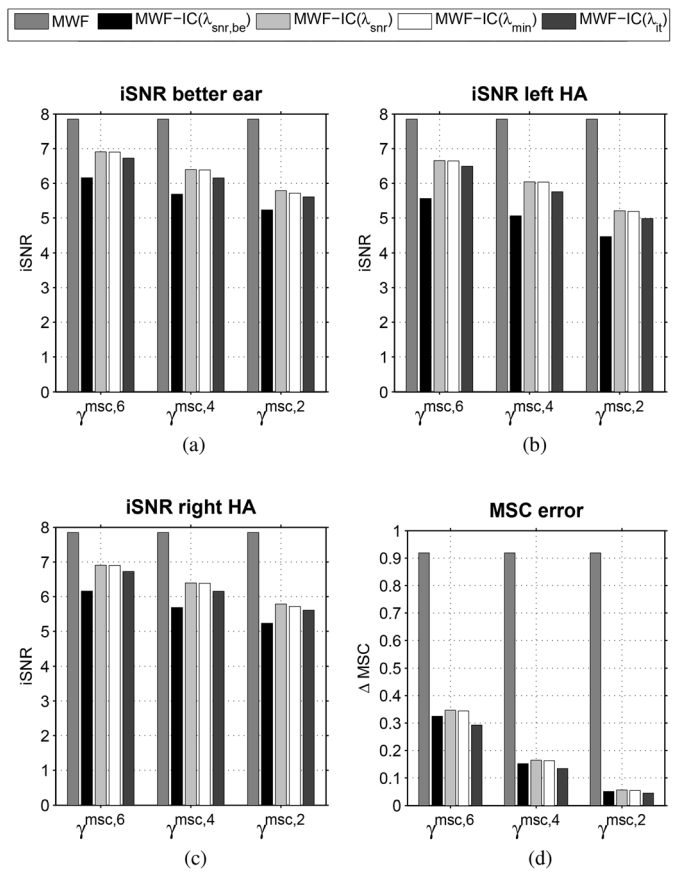


Fig. 6. Output iSNR and MSC error for the noise component for the MWF and the MWF-IC using different trade-off parameters $\lambda_{\text{snr,be}}, \lambda_{\text{snr}}, \lambda_{\min}, \text{and}\lambda_{\text{it}}$ for the MSC boundaries $\gamma^{\text{msc},6}, \gamma^{\text{msc},4}$ and $\gamma^{\text{msc},2}$. The average input iSNR in the reference microphones at the left and the right hearing aid was equal to 0 dB. (a) output iSNR (better ear) (b) output iSNR (left HA) (c) output iSNR (right HA) (d) MSC error.

Fig. 6(a)–6(c) depicts the intelligibility-weighted output SNR at the better ear (Fig. 6(a)), at the left ear (Fig. 6(b)) and at the right ear (Fig. 6(c)) for different values of the upper MSC boundary $\gamma_{\max}^{\text{msc}}$ and for different selection procedures for the trade-off parameter λ, namely $\lambda_{\text{snr,be}}$, λ_{snr}, $\lambda_{\min}$ and λ_{it} (cf. Section V-B). The better ear output iSNR for the MWF is equal to 7.8 dB. If $\lambda_{\min}$ or λ_{snr} are used, the better ear output iSNR slightly decreases to 6.9 dB ($\gamma^{\text{msc},6}$), to 6.4 dB ($\gamma^{\text{msc},4}$) and to 5.8 dB ($\gamma^{\text{msc},2}$). The results for $\lambda_{\min}$ and λ_{snr} are the same, which seems to imply that in this case the output SNR is monotonically decreasing. If λ_{it} is used, the better ear output iSNR of the MWF-IC decreases by only 0.2 dB compared to the better ear output iSNR for $\lambda_{\min}$ and λ_{snr}, showing the applicability of the iterative search method for determining a suitable trade-off parameter λ. Furthermore, when using $\lambda_{\text{snr,be}}$, the better ear output iSNR decreases by 1 dB compared to the better ear output iSNR for $\lambda_{\min}$. If the output iSNR at the left or the right hearing aid is not strictly monotonically decreasing with increasing λ, on the one hand the output iSNR for the better ear hearing aid can become slightly larger for $\lambda_{\text{snr,be}}$ than for λ_{snr}, but on the other hand the output iSNR for the other hearing aid may be much lower for $\lambda_{\text{snr,be}}$ than for λ_{snr}, possibly resulting in a decrease of the overall iSNR performance for $\lambda_{\text{snr,be}}$.

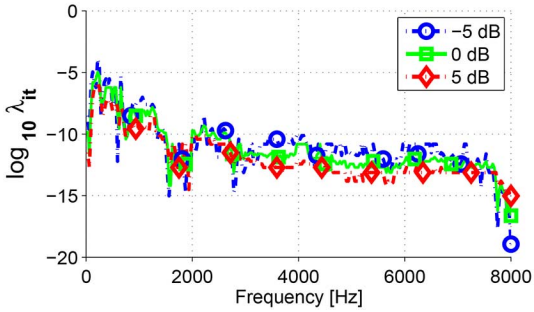


Fig. 7. Logarithmic values of λ_{it} for the MWF-IC with the MSC boundary $\gamma^{\mathrm{msc},6}$ for input iSNRs of -5 dB, 0 dB and 5 dB for the babble noise scenario.

The MSC error for different values of the upper MSC boundary $\gamma^{\mathrm{msc}}_{\mathrm{max}}$ (cf. (74)) and for different selection procedures for the trade-off parameter λ is depicted in Fig. 6(d). As expected, the MSC error is significantly reduced for the MWF-IC compared to the MWF. Furthermore, decreasing the MSC boundary $\gamma^{\mathrm{msc}}_{\mathrm{max}}$ for the MWF-IC leads to a better preservation of the output MSC of the noise component but also results in a decrease of the output iSNR, as already indicated for a specific frequency in Fig. 3. In addition, for λ_{it} and $\lambda_{\mathrm{snr,be}}$, the MSC error is always lower than for λ_{min}, which can be explained by the fact that $\lambda_{\mathrm{it}} \geq \lambda_{\mathrm{min}}$ and $\lambda_{\mathrm{snr,be}} \geq \lambda_{\mathrm{min}}$.

In conclusion, since using λ_{it} considerably decreases the computational complexity while not significantly affecting the output iSNR and the MSC error compared to λ_{snr} and λ_{min}, in the following experiments only the iterative search method (λ_{it}) will be considered.

2) Experiment 2: In the second experiment, the performance of the MWF, the MWF-ITF and the MWF-IC using λ_{it} is investigated for the babble noise and the ambient noise scenario at an average input iSNR of -5 dB, 0 dB and 5 dB. As for the first experiment, the iterative search method for the trade-off parameter λ has been initialized with $\lambda_{\mathrm{init}} = 10$. The output iSNR and the MSC error have been calculated using the speech and noise correlation matrices calculated during the 10 s speech-and-noise period.

For the *babble noise scenario*, the logarithmic values of λ_{it} for the MWF-IC using the MSC boundary $\gamma^{\mathrm{msc},6}$ and input iSNRs of -5 dB, 0 dB and 5 dB are depicted in Fig. 7. It can be generally said that for all frequencies, λ_{it} decreases with increasing input iSNR. Hence, the trade-off parameter λ_{it} depends on the input iSNR. The trade-off parameter λ_{it} for the ambient noise scenario is omitted here since it is very similar to the trade-off parameter λ_{it} for the babble noise scenario.

Fig. 8(a)–8(d) depict the iSNR gain for the left and the right hearing aid, the better ear output iSNR and the MSC error for different input iSNRs and different algorithms, i.e. the binaural MWF, MWF-ITF and MWF-IC for different MSC boundaries. As expected from the theoretical analysis in Section III, the iSNR gain at the left and the right hearing aid are the same for the MWF and the MWF-ITF. Moreover, the iSNR gain for the MWF-IC is lower than the iSNR gain for the MWF and the MWF-ITF for all MSC boundaries. While the iSNR gain of all algorithms depends on the input iSNR, the decrease in iSNR gain of the MWF-IC compared to the MWF is rather independent of the input iSNR but very much depends on the MSC boundary γ^{msc}. For example, for an input iSNR of -5 dB, the decrease in iSNR gain for the MWF-IC compared to the MWF ranges from 0.8 dB ($\gamma^{\mathrm{msc},6}$) up to 1.8 dB ($\gamma^{\mathrm{msc},2}$) in the left hearing aid and from 1 dB ($\gamma^{\mathrm{msc},6}$) up to 2.2 dB ($\gamma^{\mathrm{msc},2}$) in the right hearing aid.

As depicted in Fig. 8(d), both the MWF and the MWF-ITF yield a large MSC error. The MSC error can be significantly decreased if the MWF-IC is used, where the amount of MSC error depends on the MSC boundary γ^{msc} but not on the input iSNR. However, a lower MSC error is always associated with a decrease of the iSNR gain in the left and the right hearing aid as can be observed from Fig. 8(a)–8(b).

For the different considered algorithms, the distributions of the reliable ILD and ITD cues of the speech component, calculated according to [6], are depicted in Fig. 9. For the MWF and the MWF-IC (all MSC boundaries), it can be observed that the ILD and ITD distributions are very similar to the input distributions, with slightly larger deviations for lower input iSNRs. For an input iSNR of -5 dB (Fig. 9(a)), a slight shift for the ILD distribution of the MWF and the MWF-IC can be observed, which decreases for larger input iSNRs (Fig. 9(b)–9(c)). Although from the theoretical analysis in Section III-A a perfect preservation of the binaural cues of the speech component is expected for the MWF, this is not exactly the case due to estimation errors in the speech correlation matrix and the short STFT segment length. Furthermore, it can be observed from Fig. 9 that the impact of the IC preservation term in the MWF-IC on the binaural cues of the speech component is very small and almost independent of the desired amount of IC preservation. On the other hand, for the MWF-ITF a large deviation from the input distributions occurs, which appears to be much larger for the ILD cues than for the ITD cues. This can be explained based on the findings in Section IV-B, i.e., the output ILD cues are shifted towards negative values (cf. Fig. 2(b)), whereas the output ITD cues are shifted towards 0 for frequency bands that exhibit a positive value of the output ITF and towards $\frac{\pi}{\omega}$ for frequency bands that exhibit a negative value of the output ITF, cf. (62). Fig. 9 also shows that for an increasing iSNR the distributions of the binaural cues of the output speech component for the MWF-ITF are shifted towards the distributions of the binaural cues of the input speech component due to the SNR dependency of the output ITF of the MWF-ITF (cf. (59)). Please note that especially for the small differences between the ILD and ITD distributions for the MWF and the MWF-IC, it is not possible to make a clear statement regarding the perceived location of the speech source. Also the large errors for the MWF-ITF can not be directly mapped to a source location error. However, since the differences between the binaural cues of the input speech component and the output speech component of the MWF and the MWF-IC are very small, no impact on the perceived location of the speech source is expected, what has also been verified in informal listening tests. As expected from the ILD and ITD distributions for the MWF-ITF a perceptually unsatisfying result is obtained.

For the *ambient noise scenario*, Fig. 10(a)–10(d) depict the iSNR gain for the left and the right hearing aid, the better ear

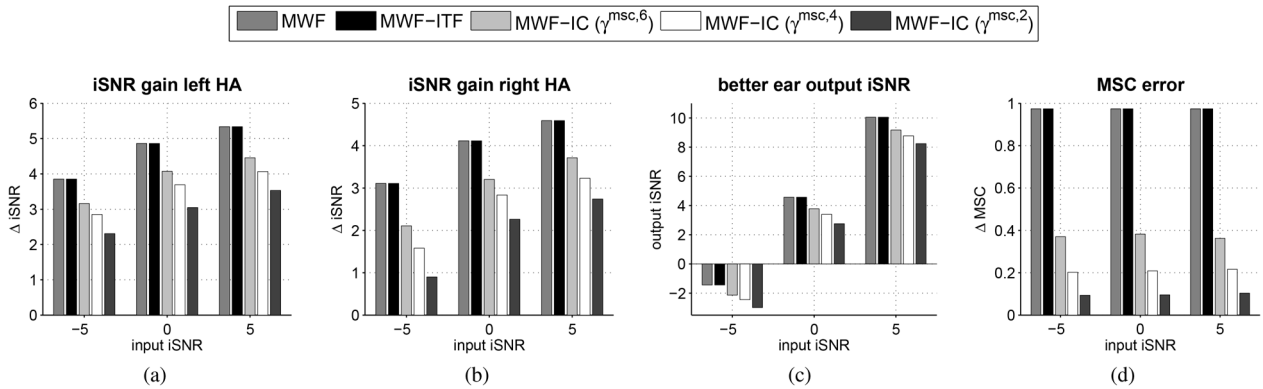


Fig. 8. iSNR gain, better ear output iSNR and MSC error of the noise component for different input iSNRs and different algorithms for the diffuse babble noise scenario.

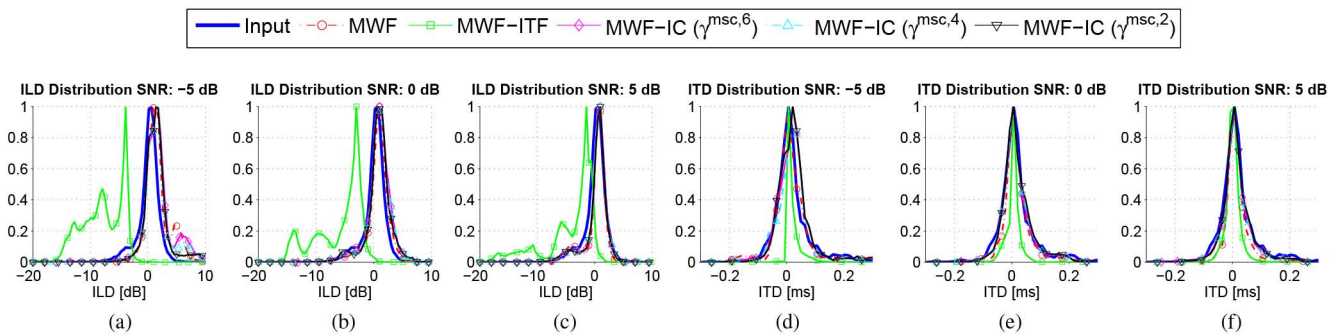


Fig. 9. Distributions of the reliable ILD and ITD cues of the speech component for different input iSNRs and different algorithms for the diffuse babble noise scenario.

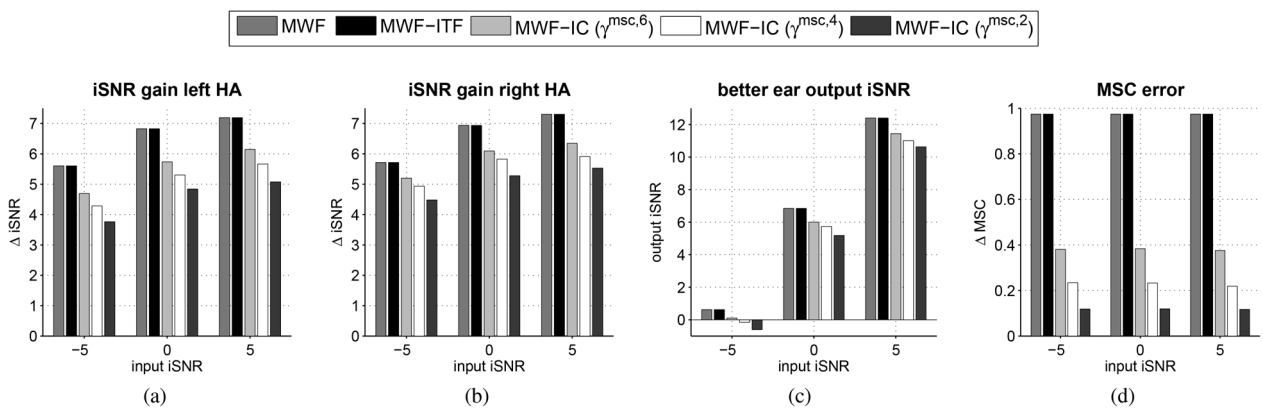


Fig. 10. iSNR gain, better ear output iSNR and MSC error of the noise component for different input iSNRs and different algorithms for the ambient noise scenario.

output iSNR and the MSC error for the noise component for different input iSNRs and for the different considered algorithms. In general, compared to the babble noise scenario, a larger iSNR gain is achieved due to the occasional presence of directional components such as interfering speakers in the ambient noise scenario. For an input iSNR of -5 dB the decrease in iSNR compared to the MWF ranges from 0.8 dB ($\gamma^{\mathrm{msc},6}$) up to 1.7 dB ($\gamma^{\mathrm{msc},2}$) in the left hearing aid and from 0.6 dB ($\gamma^{\mathrm{msc},6}$) up to 1.2 dB ($\gamma^{\mathrm{msc},2}$) in the right hearing aid. Again, the MSC error for the noise component can be significantly reduced using the MWF-IC, while the MWF-ITF exhibits the same large MSC error as the MWF. Although it was shown in Table I that the estimation errors of the correlation matrices highly depend on the input iSNR and the noise type, the MSC errors in Fig. 8(d) and 10(d) appear to be rather independent of the input iSNR and the noise type, showing that the proposed algorithm is applicable for different ranges of estimation errors for the correlation matrices.

The distributions of the reliable ILD and ITD cues for the speech component are depicted in Fig. 11. Similarly as for the babble noise scenario (cf. Fig. 9), the distributions of the binaural cues of the output speech component for the MWF and the MWF-IC are very close to the distributions of the binaural cues of the input speech component, independent of the amount of desired IC preservation. Again, the MWF-ITF shows a large deviation from the distribution of the binaural cues of the input speech component. The perceptual evaluation of the experimental results, i.e., the impact of the trade-off between IC preservation and output iSNR on speech intelligibility and spatial awareness, using subjective listening experiments, re-

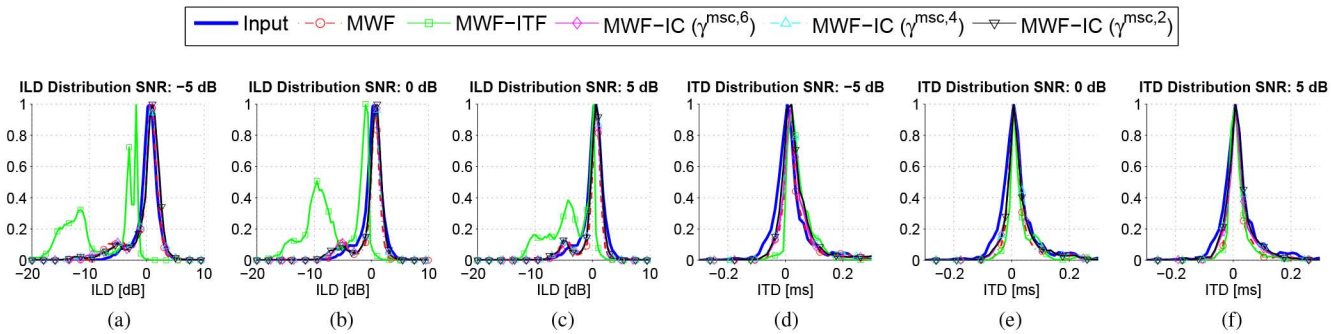


Fig. 11. Distributions of the reliable ILD and ITD cues of the speech component for different input iSNRs and different algorithms for the ambient noise scenario.

mains a topic for future research. Audio samples can be found at http://www.uni-oldenburg.de/fileadmin/user_upload/medi-physik/ag/sigproc/audio/binaural/ic.html.

VII. Conclusion

In this paper we have proposed an extension of the binaural MWF, namely the MWF-IC, aiming to preserve the interaural coherence in diffuse noise fields. The amount of IC preservation is controlled by a trade-off parameter, which has been controlled based on psychoacoustically determined MSC boundaries. Several methods for determining this trade-off parameter have been proposed and experimentally validated, showing that the iterative search method leads to a very similar result in terms of noise reduction and IC preservation as a computationally expensive exhaustive search method. In addition, a theoretical analysis of the performance of the MWF and the MWF-ITF in diffuse noise fields has been presented and validated using simulations in different diffuse noise scenarios. Furthermore, we have shown that the MWF-IC yields a better preservation of the output IC of the noise component compared to the MWF and the MWF-ITF without significantly distorting the binaural cues of the speech component but at the expense of a degraded noise reduction performance.

Appendix A

In this appendix, we derive the gradient for the cost function $J_{\mathrm{MWF-IC}}(\mathbf{W})$ in (65). We first decompose the $2M$-dimensional complex-valued vector $\mathbf{W}$ into its real and imaginary parts, which are denoted by $\mathbf{W}_R$ and $\mathbf{W}_I$. We define the $4M$-dimensional real-valued weight vector $\widetilde{\mathbf{W}}$ as

$$\widetilde{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_R \\ \mathbf{W}_I \end{bmatrix} = \begin{bmatrix} \mathbf{W}_{0R} \\ \mathbf{W}_{1R} \\ \mathbf{W}_{0I} \\ \mathbf{W}_{1I} \end{bmatrix}. \tag{85}$$

The cost function in (65) can now be written as

$$J_{MWF-IC}(\widetilde{\mathbf{W}}) = J_{MWF}(\widetilde{\mathbf{W}}) + \lambda J_{IC}(\widetilde{\mathbf{W}}). \tag{86}$$

Using (32), the cost function $J_{MWF}(\widetilde{\mathbf{W}})$ can be written as

$$J_{MWF}(\widetilde{\mathbf{W}}) = \left(\mathbf{W}^H\mathbf{R}\mathbf{W}\right) - 2\left(\mathbf{W}^H\mathbf{r}_x\right)_R + P, \tag{87}$$

$$= \widetilde{\mathbf{W}}^T\widetilde{\mathbf{R}}_R\widetilde{\mathbf{W}} - 2\widetilde{\mathbf{W}}^T\widetilde{\mathbf{r}}_{x,R} + P, \tag{88}$$

with $P = P_s|A_0|^2 + P_s|A_1|^2$ and

$$\widetilde{\mathbf{R}}_R = \begin{bmatrix} \mathbf{R}_R & -\mathbf{R}_I \\ \mathbf{R}_I & \mathbf{R}_R \end{bmatrix}, \quad \widetilde{\mathbf{r}}_{x,R} = \begin{bmatrix} \mathbf{r}_{x,R} \\ \mathbf{r}_{x,I} \end{bmatrix}. \tag{89}$$

The gradient of $J_{MWF}(\widetilde{\mathbf{W}})$ in (88) is then equal to

$$\frac{\partial J_{MWF}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}} = \left(\widetilde{\mathbf{R}}_R + \widetilde{\mathbf{R}}_R^T\right)\widetilde{\mathbf{W}} - 2\widetilde{\mathbf{r}}_{x,R}. \tag{90}$$

The cost function $J_{IC}(\mathbf{W})$ in (63) can be written as

$$J_{IC} = \left[\frac{\left(\mathbf{W}^H\mathbf{R}_v^{01}\mathbf{W}\right)_R}{\sqrt{\left(\mathbf{W}^H\mathbf{R}_v^{00}\mathbf{W}\right)\left(\mathbf{W}^H\mathbf{R}_v^{11}\mathbf{W}\right)}} - \alpha_R\right]^2 + \left[\frac{\left(\mathbf{W}^H\mathbf{R}_v^{01}\mathbf{W}\right)_I}{\sqrt{\left(\mathbf{W}^H\mathbf{R}_v^{00}\mathbf{W}\right)\left(\mathbf{W}^H\mathbf{R}_v^{11}\mathbf{W}\right)}} - \alpha_I\right]^2, \tag{91}$$

with $\alpha = IC_v^{des}$ and

$$\mathbf{R}_v^{01} = \begin{bmatrix} \mathbf{0}_{2M} & \mathbf{R}_v \\ \mathbf{0}_{2M} & \mathbf{0}_{2M} \end{bmatrix}, \quad \mathbf{R}_v^{00} = \begin{bmatrix} \mathbf{R}_v & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{0}_{2M} \end{bmatrix}, \tag{92}$$

$$\mathbf{R}_v^{11} = \begin{bmatrix} \mathbf{0}_{2M} & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{R}_v \end{bmatrix}. \tag{93}$$

Hence, using (85) and (91) the cost function $J_{IC}(\widetilde{\mathbf{W}})$ can be written as

$$J_{IC}(\widetilde{\mathbf{W}}) = \left[\frac{\widetilde{\mathbf{W}}^T\widetilde{\mathbf{R}}_{v,R}^{01}\widetilde{\mathbf{W}}}{\sqrt{\left(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{R}}_{v,R}^{00}\widetilde{\mathbf{W}}\right)\left(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{R}}_{v,R}^{11}\widetilde{\mathbf{W}}\right)}} - \alpha_R\right]^2 + \left[\frac{\widetilde{\mathbf{W}}^T\widetilde{\mathbf{R}}_{v,I}^{01}\widetilde{\mathbf{W}}}{\sqrt{\left(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{R}}_{v,R}^{00}\widetilde{\mathbf{W}}\right)\left(\widetilde{\mathbf{W}}^T\widetilde{\mathbf{R}}_{v,R}^{11}\widetilde{\mathbf{W}}\right)}} - \alpha_I\right]^2, \tag{94}$$

where the first part is denoted as $J_{IC,R}(\widetilde{\mathbf{W}})$ and the second part is denoted as $J_{IC,I}(\widetilde{\mathbf{W}})$ and

$$\widetilde{\mathbf{R}}_{v,R}^{01} = \begin{bmatrix} \mathbf{R}_{v,R}^{01} & -\mathbf{R}_{v,I}^{01} \\ \mathbf{R}_{v,I}^{01} & \mathbf{R}_{v,R}^{01} \end{bmatrix}, \ \widetilde{\mathbf{R}}_{v,R}^{00} = \begin{bmatrix} \mathbf{R}_{v,R}^{00} & -\mathbf{R}_{v,I}^{00} \\ \mathbf{R}_{v,I}^{00} & \mathbf{R}_{v,R}^{00} \end{bmatrix},$$

$$\widetilde{\mathbf{R}}_{v,I}^{01} = \begin{bmatrix} \mathbf{R}_{v,I}^{01} & \mathbf{R}_{v,R}^{01} \\ -\mathbf{R}_{v,R}^{01} & \mathbf{R}_{v,I}^{01} \end{bmatrix}, \ \widetilde{\mathbf{R}}_{v,R}^{11} = \begin{bmatrix} \mathbf{R}_{v,R}^{11} & -\mathbf{R}_{v,I}^{11} \\ \mathbf{R}_{v,I}^{11} & \mathbf{R}_{v,R}^{11} \end{bmatrix}.$$

To simplify the notation, we define

$$\widetilde{\mathbf{A}} = \widetilde{\mathbf{R}}_{v,R}^{01}, \quad \widetilde{\mathbf{B}} = \widetilde{\mathbf{R}}_{v,R}^{00}, \quad \widetilde{\mathbf{C}} = \widetilde{\mathbf{R}}_{v,R}^{11}, \tag{95}$$

and

$$D = \frac{(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{A}} \widetilde{\mathbf{W}})}{(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{B}} \widetilde{\mathbf{W}})(\widetilde{\mathbf{W}}^T \widetilde{\mathbf{C}} \widetilde{\mathbf{W}})}. \tag{96}$$

The gradient of $J_{IC,R}(\widetilde{\mathbf{W}})$ is then equal to

$$\frac{\partial J_{IC,R}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}} = D\,(2D - \alpha_R) \times \left[\frac{\left(\widetilde{\mathbf{A}} + \widetilde{\mathbf{A}}^T\right)\widetilde{\mathbf{W}}}{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{A}} \widetilde{\mathbf{W}}} - \frac{1}{2} \frac{\left(\widetilde{\mathbf{B}} + \widetilde{\mathbf{B}}^T\right)\widetilde{\mathbf{W}}}{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{B}} \widetilde{\mathbf{W}}} - \frac{1}{2} \frac{\left(\widetilde{\mathbf{C}} + \widetilde{\mathbf{C}}^T\right)\widetilde{\mathbf{W}}}{\widetilde{\mathbf{W}}^T \widetilde{\mathbf{C}} \widetilde{\mathbf{W}}} \right]. \tag{97}$$

The gradient of $J_{IC,I}(\widetilde{\mathbf{W}})$ can be computed similarly as in (97) by setting $\widetilde{\mathbf{A}} = \widetilde{\mathbf{R}}_{v,I}^{01}$ and $\alpha_R = \alpha_I$. The gradient of the overall cost function can then be calculated by combining (90) and (97), i.e.,

$$\frac{\partial J_{MWF-IC}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}} = \frac{\partial J_{MWF}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}} + \lambda \left(\frac{\partial J_{IC,R}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}} + \frac{\partial J_{IC,I}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}} \right).$$

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] 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.

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

[4] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1648–1661, 1992.

[5] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *J. Acoust. Soc. Amer.*, vol. 116, no. 5, pp. 3075–3089, 2004.

[6] 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.

[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. Amer.*, vol. 74, no. 6, pp. 1726–1733, 1983.

[8] M. Vorländer, *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. New York, NY, USA: Springer, 2008.

[9] T. Hidaka, L. L. Beranek, and T. Okano, "Interaural cross-correlation, lateral fraction, and low- and high-frequency sound levels as measures of acoustical quality in concert halls," *J. Acoust. Soc. Amer.*, vol. 98, no. 2, pp. 988–1007, 1995.

[10] 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.

[11] 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.

[12] M. Lavandier and J. F. Culling, "Prediction of binaural speech intelligibility against noise in rooms," *J. Acoust. Soc. Amer.*, vol. 127, no. 1, pp. 387–399, 2010.

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

[14] 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.

[15] M. Dörbecker and S. Ernst, "Combination of two–channel spectral subtraction and adaptive wiener post–filtering for noise reduction and dereverberation," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Trieste, Italy, Sep. 1996, pp. 995–998.

[16] 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.

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

[18] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures," in *Proc. WASPAA*, New Paltz, NY, USA, Oct. 2007, pp. 315–318.

[19] G. Grimm, V. Hohmann, and B. Kollmeier, "Increase and subjective evaluation of feedback stability in hearing aids by a binaural coherence-based noise reduction scheme," *Proc. IEEE Trans. Audio, Speech Lang.*, vol. 17, no. 7, pp. 1408–1419, Sep. 2009.

[20] K. Reindl, Y. Zheng, and W. Kellermann, "Analysis of two generic wiener filtering concepts for binaural speech enhancement in hearing aids," in *Proc. EUSIPCO*, Aalborg, Denmark, Aug. 2010, pp. 989–993.

[21] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous binaural target PSD estimation for hearing aid noise reduction in complex acoustic environments," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 4, pp. 1141–1154, Apr. 2011.

[22] R. C. Hendriks, T. Gerkmann, and J. Jensen, *DFT-Domain Based Single-Microphone Noise Reduction for Speech Enhancement - A Survey of the State of the Art*, ser. Synthesis Lectures on Speech and Audio Processing. San Rafael, CA, USA: Morgan & Claypool, 2013.

[23] 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.

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

[25] 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. ICASSP*, Honolulu, HI, USA, Apr. 2007, pp. 565–568.

[26] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in *Proc. Int. Workshop Acoust. Signal Enhance. (IWAENC)*, Aachen, Germany, Sep. 2012, pp. 117–120.

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

[28] D. Marquardt, V. Hohmann, and S. Doclo, "Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. ICASSP*, Vancouver, BC, Canada, May 2013, pp. 8648–8652.

[29] D. Marquardt, V. Hohmann, and S. Doclo, "Perceptually motivated coherence preservation in multi-channel wiener filtering based noise reduction for binaural hearing aids," in *Proc. ICASSP*, Florence, Italy, May 2014, pp. 3660–3664.

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

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

[32] S. Braun, M. Torcoli, D. Marquardt, E. Habets, and S. Doclo, "Multichannel dereverberation for hearing aids with interaural coherence preservation," in *Proc. 14th Int. Workshop Acoust. Signal Enhance. (IWAENC)*, Sep. 2014, pp. 124–128.

[33] 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.

[34] M. Jeub, M. Dorbecker, 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.
[35] 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.
[36] Mathworks. Matlab Optimization Toolbox User's Guide [Online]. Available: http://www.mathworks.com/help/pdf_doc/optim/optim_tb.pdf
[37] R. Fletcher, *Practical Methods of Optimization*. New York, NY, USA: Wiley-Interscience, 1987.
[38] K. J. Gabriel and H. S. Colburn, "Interaural correlation discrimination: I. bandwidth and level dependence," *J. Acoust. Soc. Amer.*, vol. 69, no. 5, pp. 1394–1401, May 1981.
[39] J. F. Culling, H. S. Colburn, and M. Spurchise, "Interaural correlation sensitivity," *J. Acoust. Soc. Amer.*, vol. 110, no. 2, pp. 1020–1029, Aug. 2001.
[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. Adv. Signal Process.*, vol. 2009, p. 10, 2009.
[41] 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.
[42] 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.
[43] 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.
[44] B. A. Edmonds and J. F. Culling, "The spatial unmasking of speech: Evidence for better-ear listening," *J. Acoust. Soc. Amer.*, vol. 120, no. 3, pp. 1539–1545, 2006.

Daniel Marquardt (S'12) received the Dipl.-Ing. degree in media technology with focus on audio-visual technology from Ilmenau University of Technology, Germany, in 2010. From 2009 to 2010, he was with Nuance Communications, Inc and Fraunhofer IDMT where he worked in the field of voice activity detection and binaural acoustics. Since 2010 he has been a Ph.D. student in the Signal Processing Group at the University of Oldenburg, Germany. His research interests are in the area of signal processing for binaural hearing aids.

Volker Hohmann received the Physics degree (Dipl. -Phys.) and the doctorate degree in Physics (Dr. rer. nat.) from the University of Göttingen, Germany, in 1989 and 1993, respectively. Since 1993, he has been a Faculty Member of the Physics Institute, University of Oldenburg, Germany, and member of the Medical Physics Group (Prof. B. Kollmeier). He has been active in teaching activities in physics for undergraduate and graduate courses and has research expertise in acoustics and digital signal processing with applications to signal processing in speech processing devices, e.g., hearing aids. He is a Consultant with the Hörzentrum Oldenburg GmbH. He was a Guest Researcher at Boston University, Boston, MA, (Prof. Dr. Colburn) in 2000 and at the Technical University of Catalonia, Barcelona, Spain in 2008. Prof. Hohmann received the Lothar-Cremer price of the German acoustical society (DEGA) in 2008 and the German President's Award for Technology and Innovation in 2012.

Simon Doclo (S'95–M'03–SM'13) received the M.Sc. degree in electrical engineering and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Belgium, in 1997 and 2003. From 2003 to 2007 he was a Postdoctoral Fellow with the Research Foundation Flanders at the Electrical Engineering Department (Katholieke Universiteit Leuven) and the Adaptive Systems Laboratory (McMaster University, Canada). From 2007 to 2009 he was a Principal Scientist with NXP Semiconductors at the Sound and Acoustics Group in Leuven, Belgium. Since 2009 he has been a Full Professor at the University of Oldenburg, Germany, and scientific advisor for the project group Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology. His research activities center around signal processing for acoustical applications, more specifically microphone array processing, active noise control, acoustic sensor networks and hearing aid processing. Prof. Doclo received the Master Thesis Award of the Royal Flemish Society of Engineers in 1997 (with Erik De Clippel), the Best Student Paper Award at the International Workshop on Acoustic Echo and Noise Control in 2001, the EURASIP Signal Processing Best Paper Award in 2003 (with Marc Moonen) and the IEEE Signal Processing Society 2008 Best Paper Award (with Jingdong Chen, Jacob Benesty, Arden Huang). He was member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing (2008–2013) and Technical Program Chair for the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in 2013. Prof. Doclo has served as guest editor for several special issues (IEEE Signal Processing Magazine, Elsevier Signal Processing) and is associate editor for IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and *EURASIP Journal on Advances in Signal Processing*.