Comparison of RTF Estimation Methods between a Head-Mounted Binaural Hearing Device and an External Microphone

Nico Gößling, Daniel Marquardt, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics and
Cluster of Excellence Hearing4All, Oldenburg, Germany

{nico.goessling,daniel.marquardt,simon.doclo}@uni-oldenburg.de

Abstract

Besides noise reduction, an important objective of a binaural speech enhancement algorithm is the preservation of the binaural cues of both the desired speech source as well as the undesired noise in order to preserve the spatial impression of the acoustic scene for the listener. Recently, it has been shown for the binaural MVDR beamformer with partial noise estimation (MVDR-N) that by combining head-mounted hearing devices with an external microphone it is possible to improve the noise reduction performance while achieving the same binaural cue preservation. While the relative positions of the head-mounted microphones can be assumed to be stationary this assumption does not hold for the external microphone, which can change its relative position due to head movements or direct movement of the listener or the external microphone. In this paper, we compare the influence of different methods for estimating the relative transfer functions of the desired speech source between the head-mounted microphones and the external microphone on the noise reduction and binaural cue preservation performance of the binaural MVDR-N beamformer.

Index Terms: binaural cues, noise reduction, external microphone, interaural coherence, relative transfer functions

1. Introduction

Noise reduction algorithms for head-mounted hearing devices (e.g., hearing aids) are crucial to improve speech quality and intelligibility in background noise. Binaural devices, consisting of one or more microphones on each side of the head of the listener, are able to exploit not only spectral but also spatial information on both sides of head [1–3]. Besides noise reduction, preserving the binaural cues of all present sound sources is an important task of a binaural noise reduction algorithm in order to ensure that the listener's spatial impression is not distorted by the algorithm.

For a single desired speech source, the binaural multi-channel Wiener filter (MWF) [2, 4] has been shown to preserve the binaural cues of the desired speech source. However, it typically distorts the binaural cues of the noise, such that the residual noise is perceived as coming from the same direction as the desired speech source which is obviously undesired. As an extension, the binaural MWF with partial noise estimation (MWF-N) has been proposed [2, 4, 5], which aims at preserving the speech component and a scaled version of the noise component in the reference microphones of the left and the right hearing device. It has been shown that the mixing parameter in the binaural MWF-N allows to trade off noise reduction and binaural cue preservation performance of the noise component [4].

In this paper we consider the binaural minimum variance distortionless response (MVDR) beamformer with partial noise estimation (MVDR-N) [2, 4–6], which can be considered as a special case of the binaural MWF-N only performing spatial processing. Recently, the use of one or more external microphones (eMics) in combination with head-mounted hearing devices (HHDs) have been explored [7–13]. It has been shown that using an eMic can increase both noise reduction and binaural cue preservation performance, depending on the position of the eMic [10, 12].

To implement the binaural MVDR beamformer, an estimate of the relative transfer functions (RTFs) of the desired speech source between all microphones and the reference microphones on both HHDs are required. Instead of using *reverberant* RTFs, one can also use *anechoic* RTFs. When an estimate of the direction-of-arrival (DOA) of the desired speech source is available these anechoic RTFs can be easily constructed for the head-mounted microphones, e.g., based on measurements or head models. However, even when the DOA of the desired speech source (relative to head) is known, this can not be used to compute the (anechoic or reverberant) RTF between the reference microphones and the eMic, since the position of the eMic is not known. Hence, the (anechoic or reverberant) RTF needs to be estimated from the microphone signals.

In this paper, we investigate the influence of three different RTF estimation methods [14–17] on the noise reduction and binaural cue preservation performance of the binaural MVDR-N beamformer for a scenario with one desired speech source surrounded by diffuse multi-talker noise in a reverberant environment. As will be seen, the so-called covariance whitening [14, 15, 17] outperforms the others in terms of noise reduction and binaural cue preservation performance.

2. Configuration and notation

2.1. Signal model

Consider the multiple-input binaural-output (MIBO) system depicted in Fig. 1, consisting of a HHD with M_L microphones on the left side of the head, a HHD with M_R microphones on the right side of the head and an additional eMic, located somewhere else in the room at an unknown position. The m-th microphone signal in the left HHD $Y_{L,m}(\omega)$ can be written in the frequency-domain as

$$Y_{L,m}(\omega) = X_{L,m}(\omega) + N_{L,m}(\omega), \quad m = 1, \ldots, M_L, \quad (1)$$

with $X_{L,m}(\omega)$ the speech component and $N_{L,m}(\omega)$ the noise component. The m-th microphone signal in the right HHD $Y_{R,m}(\omega)$ can be written similarly. The eMic signal $Y_e(\omega)$ can

This work was supported in part by the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG), and by the joint Lower Saxony-Israeli Project ATHENA.

be written as

$$Y_e(\omega) = X_e(\omega) + N_e(\omega), \tag{2}$$

with $X_e(\omega)$ the speech component and $N_e(\omega)$ the noise component in the eMic signal. For conciseness, we will omit the frequency variable ω in the remainder of the paper whenever possible. All microphone signals can be stacked in an M-dimensional vector, with $M = M_L + M_R + 1$, as

$$\mathbf{y} = \left[Y_{L,1} \ \ldots \ Y_{L,M_L} \ Y_{R,1} \ \ldots \ Y_{R,M_R} \ Y_e\right]^T, \tag{3}$$

which can be written as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{4}$$

where the vectors $\mathbf{x}$ and $\mathbf{n}$ are defined similarly as (3).

For a single desired speech source the speech vector $\mathbf{x}$ is given by

$$\mathbf{x} = \mathbf{a}S, \tag{5}$$

where the vector $\mathbf{a}$ contains the acoustic transfer functions (ATFs) between the desired speech source and all microphones and S is the (dry) speech signal. Please note that in the time-domain the vector $\mathbf{a}$ corresponds to the room impulse responses (RIRs) between the desired speech source and all microphones and hence includes reverberation.

Without loss of generality, we define the first microphone of both HHDs as the reference microphones. For ease of notation, the reference microphone signals $Y_{L,1}$ and $Y_{R,1}$ are further denoted as Y_L and Y_R and can be written as

$$Y_L = \mathbf{e}_L^T\mathbf{y}, \quad Y_R = \mathbf{e}_R^T\mathbf{y}, \tag{6}$$

where $\mathbf{e}_L$ and $\mathbf{e}_R$ denote M-dimensional zero vectors with $\mathbf{e}_L(1) = 1$ and $\mathbf{e}_R(M_L + 1) = 1$. Similarly, the eMic signal can be written as $Y_e = \mathbf{e}_e^T\mathbf{y}$, with $\mathbf{e}_e = [0 \ldots 1]^T$. Using (6), the reference microphone signals can be written as

$$Y_L = \underbrace{A_L S}_{X_L} + N_L, \quad Y_R = \underbrace{A_R S}_{X_R} + N_R, \tag{7}$$

where $A_L = \mathbf{e}_L^T\mathbf{a}$ and $A_R = \mathbf{e}_R^T\mathbf{a}$ denote the ATFs between the reference microphones and the desired speech source. The anechoic ATFs (not including reverberation) are denoted as $\bar{A}_L$ and $\bar{A}_R$. The RTF vectors for the left and the right HHD, relating the ATF vector $\mathbf{a}$ to the reference microphones [15, 16], are defined as

$$\mathbf{h}_L = \frac{\mathbf{a}}{A_L}, \quad \mathbf{h}_R = \frac{\mathbf{a}}{A_R}. \tag{8}$$

The speech and noise correlation matrices are given by

$$\mathbf{R}_x = \mathcal{E}\left\{\mathbf{x}\mathbf{x}^H\right\} = \Phi_s\mathbf{a}\mathbf{a}^H, \tag{9}$$

$$\mathbf{R}_n = \mathcal{E}\left\{\mathbf{n}\mathbf{n}^H\right\}, \tag{10}$$

with $\mathcal{E}\{\cdot\}$ the expectation operator, ^{H} the conjugate transpose and $\Phi_s = \mathcal{E}\left\{|S|^2\right\}$ the power spectral density (PSD) of the speech signal. The noise correlation matrix is assumed to be full rank and hence invertible. By assuming statistical independence between $\mathbf{x}$ and $\mathbf{n}$, the correlation matrix of the microphone signals can be written as

$$\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_n. \tag{11}$$

The (binaural) output signals of the left and the right HHD are obtained by filtering *all* microphone signals, including the external microphone signal, with the complex-valued filter vectors $\mathbf{w}_L$ and $\mathbf{w}_R$, respectively, i.e.,

$$Z_L = \mathbf{w}_L^H\mathbf{y}, \quad Z_R = \mathbf{w}_R^H\mathbf{y}. \tag{12}$$

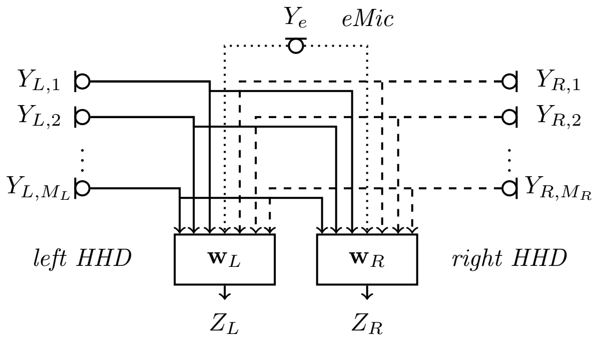

Figure 1: *MIBO system consisting of two head-mounted hearing devices and an external microphone*

The speech component in the output signals is given by $Z_{x,L}$ and $Z_{x,R}$.

2.2. Binaural cues

In addition to monaural cues, binaural cues are used by the listener to localize sound sources and to get a sense of the surrounding sound field [18, 19]. For coherent (directional) sound sources the most descriptive binaural cues are the interaural level difference (ILD) and the interaural time difference (ITD). The interaural coherence (IC) is important for source localization in multi-source and reverberant environments since it determines the reliability of the ILD and ITD cues [19, 20].

The input interaural transfer function (ITF) of the speech component is defined as

$$\mathrm{ITF}_x^{\mathrm{in}} = \frac{\mathcal{E}\left\{X_L X_R^*\right\}}{\mathcal{E}\left\{|X_R|^2\right\}} = \frac{\mathbf{e}_L^T\mathbf{R}_x\mathbf{e}_R}{\mathbf{e}_R^T\mathbf{R}_x\mathbf{e}_R}. \tag{13}$$

The output ITF of the speech component is similarly defined as

$$\mathrm{ITF}_x^{\mathrm{out}} = \frac{\mathcal{E}\left\{Z_{x,L} Z_{x,R}^*\right\}}{\mathcal{E}\left\{|Z_{x,R}|^2\right\}} = \frac{\mathbf{w}_L^H\mathbf{R}_x\mathbf{w}_R}{\mathbf{w}_R^H\mathbf{R}_x\mathbf{w}_R}. \tag{14}$$

The input ILD of the speech component is defined as the power ratio of the speech component in the left and the right HHD [4], i.e.,

$$\mathrm{ILD}_x^{\mathrm{in}} = \frac{\mathcal{E}\left\{|X_L|^2\right\}}{\mathcal{E}\left\{|X_R|^2\right\}} = \frac{\mathbf{e}_L^T\mathbf{R}_x\mathbf{e}_L}{\mathbf{e}_R^T\mathbf{R}_x\mathbf{e}_R}. \tag{15}$$

The output ILD of the speech component is similarly defined as

$$\mathrm{ILD}_x^{\mathrm{out}} = \frac{\mathcal{E}\left\{|Z_{x,L}|^2\right\}}{\mathcal{E}\left\{|Z_{x,R}|^2\right\}} = \frac{\mathbf{w}_L^H\mathbf{R}_x\mathbf{w}_L}{\mathbf{w}_R^H\mathbf{R}_x\mathbf{w}_R}. \tag{16}$$

The ITD can be calculated from the ITF as [4]

$$\mathrm{ITD} = \frac{\angle\mathrm{ITF}}{\omega}, \tag{17}$$

with $\angle$ denoting the phase. The input noise IC is defined as

$$\mathrm{IC}_n^{\mathrm{in}} = \frac{\mathbf{e}_L^T\mathbf{R}_n\mathbf{e}_R}{\sqrt{(\mathbf{e}_L^T\mathbf{R}_n\mathbf{e}_L)(\mathbf{e}_R^T\mathbf{R}_n\mathbf{e}_R)}}. \tag{18}$$

The output noise IC is similarly defined as

$$\mathrm{IC}_n^{\mathrm{out}} = \frac{\mathbf{w}_L^H\mathbf{R}_n\mathbf{w}_R}{\sqrt{(\mathbf{w}_L^H\mathbf{R}_n\mathbf{w}_L)(\mathbf{w}_R^H\mathbf{R}_n\mathbf{w}_R)}}. \tag{19}$$

The (real-valued) magnitude-squared coherence (MSC) is defined as $\mathrm{MSC} = |\mathrm{IC}|^2$.

3. Binaural noise reduction

In this section we introduce a binaural noise reduction approach that uses all microphones to spatially filter the microphone inputs. The binaural MVDR-N beamformer [2, 4–6] minimizes the output noise PSD while preserving the speech component in the reference microphone signals (and hence the binaural cues of the speech component) and a scaled version of the noise component in the reference microphone signals. The constraint optimization problem for the left filter can be formulated as

$$\min_{\mathbf{w}_L} \mathcal{E}\left\{|\mathbf{w}_L^H\mathbf{n} - \eta N_L|^2\right\} \quad \text{s.t.} \quad \mathbf{w}_L^H\mathbf{a} = A_L. \tag{20}$$

The solution for the left filter is given by [6, 21]

$$\mathbf{w}_{\eta,L} = (1-\eta)\overbrace{\frac{\mathbf{R}_n^{-1}\mathbf{a}}{\mathbf{a}^H\mathbf{R}_n^{-1}\mathbf{a}}A_L^*}^{\mathbf{w}_{0,L}} + \eta\mathbf{e}_L, \tag{21}$$

$$= (1-\eta)\frac{\mathbf{R}_n^{-1}\mathbf{h}_L}{\mathbf{h}_L^H\mathbf{R}_n^{-1}\mathbf{h}_L} + \eta\mathbf{e}_L, \tag{22}$$

with $0 \leq \eta \leq 1$ a real-valued mixing parameter. The solution for the right filter is similar to (22) by substituting R for L. Using (22) in (12), the output of the binaural MVDR-N beamformer can be interpreted as a mixture between the binaural MVDR beamformer output (scaled with $1-\eta$) and the (noisy) reference microphone signal (scaled with η).

For $\eta = 0$ the binaural MVDR-N beamformer is equal to the binaural MVDR beamformer $\mathbf{w}_{0,L}$ [2, 3, 22] and hence preserves the ILD and ITD cues of the desired speech source [4]. However, it has been shown in [6] that for the binaural MVDR beamformer the output noise MSC is equal to 1 and hence the surrounding noise field is perceived as coming from the same direction as the desired speech source. For $\eta = 1$ the binaural MVDR-N beamformer output is equal to the reference microphone signals in (6) and hence preserves the binaural cues of both the desired speech source and the noise component, although no noise reduction is achieved. Hence, the binaural MVDR-N beamformer trades off noise reduction against binaural cue preservation of the noise component using the mixing parameter η.

Since accurately estimating the ATF vector $\mathbf{a}$ is known to be difficult [23], several methods for estimating the RTF vectors $\mathbf{h}_L$ and $\mathbf{h}_R$ have been proposed [14–17] and hence the usage of (22) is preferred. If all microphone positions are known and a reliable DOA estimation is available, one can also use measured [24] or simulated [25] *anechoic* RTF vectors. While this is a reasonable (and robust) approach when only using the head-mounted microphones, the exact position of the eMic is usually not known. Hence, at least for the eMic, other methods, e.g., estimated RTFs between the reference microphones and the eMic need to be considered.

Due to robustness, we use anechoic RTFs for the head-mounted microphones (assuming the DOA θ to be known) and estimated RTFs only for the eMic, i.e.,

$$\boxed{\tilde{\mathbf{h}}_L = \begin{bmatrix}\bar{\mathbf{h}}_L(\theta)\\ H_{e,L}\end{bmatrix}, \quad \tilde{\mathbf{h}}_R = \begin{bmatrix}\bar{\mathbf{h}}_R(\theta)\\ H_{e,R}\end{bmatrix}} \tag{23}$$

where $\bar{\mathbf{h}}_L(\theta)$ and $\bar{\mathbf{h}}_R(\theta)$ denote the M_L- and M_R-dimensional anechoic (measured or simulated) RTF vectors which depend on the DOA θ for the left and the right HHD, respectively, and $H_{e,L}$ and $H_{e,R}$ denote the estimated (anechoic or reverberant) RTFs between the HHD reference microphones and the eMic. The construction of the RTF vectors is schematically depicted in Fig. 2.

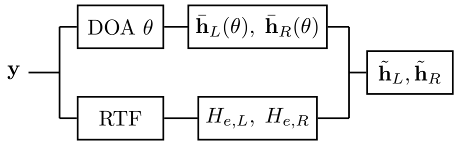
y
DOA θ
$\bar{\mathbf{h}}_L(\theta),\ \bar{\mathbf{h}}_R(\theta)$
RTF
$H_{e,L},\ H_{e,R}$
$\tilde{\mathbf{h}}_L, \tilde{\mathbf{h}}_R$

Figure 2: *Proposed construction of the RTF vectors*

By using anechoic RTFs for both the head-mounted microphones and the eMic the RTF vectors are connected by a simple factor, i.e., $\tilde{\mathbf{h}}_L = \tilde{\mathbf{h}}_R\frac{\bar{A}_L}{\bar{A}_R}$ and hence are parallel. This leads to the aforementioned mapping of the noise component to the position of the desired speech source and hence the output noise MSC being equal to 1. By mixing anechoic and reverberant RTFs, i.e., estimating reverberant RTFs for the eMic, the RTF vectors are not parallel, which leads to partial cue preservation of the noise component even when the mixing parameter η is set to 0 as will be seen in the experimental results in Section 5.

4. RTF estimation methods

In this section we describe three different methods to estimate the RTFs $H_{e,L}$ and $H_{e,R}$ between the head-mounted reference microphones and the eMic which are then used in (23). Although only the estimators for $H_{e,L}$ are discussed, the estimators for $H_{e,R}$ can again simply be obtained by substituting R for L. Using the speech correlation matrix in (9), the RTF between the left reference microphone and the eMic is given by

$$H_{e,L} = \frac{\mathbf{e}_e^T\mathbf{R}_x\mathbf{e}_L}{\mathbf{e}_L^T\mathbf{R}_x\mathbf{e}_L} = \frac{A_e}{A_L}. \tag{24}$$

4.1. Biased approach

Assuming a reasonable large SNR, the speech correlation matrix in (24) can simply be replaced by the (noisy) correlation matrix of the microphone signals $\mathbf{R}_y$ in (11), leading to the biased estimator

$$\boxed{H_{e,L}^{\mathrm{b}} = \frac{\mathbf{e}_e^T\mathbf{R}_y\mathbf{e}_L}{\mathbf{e}_L^T\mathbf{R}_y\mathbf{e}_L} = \frac{\mathcal{E}\left\{Y_eY_L^*\right\}}{\mathcal{E}\left\{|Y_L|^2\right\}}} \tag{25}$$

Generally, by using the biased estimator in (25) to estimate $H_{e,L}^{\mathrm{b}}$ and $H_{e,R}^{\mathrm{b}}$, the RTF vectors in (23) are not parallel.

4.2. MVDR pre-processed RTF estimation

An alternative approach to estimate the RTFs was proposed in [13], where it was proposed to pre-process the head-mounted microphones using an MVDR beamformer. The binaural MVDR beamformer only using the HHDs can be written in terms of the anechoic RTFs vectors $\bar{\mathbf{h}}_L(\theta)$ and $\bar{\mathbf{h}}_R(\theta)$ as

$$\mathbf{w}_{\mathrm{H},L} = \begin{bmatrix}\frac{\mathbf{R}_{n,\mathrm{H}}^{-1}\bar{\mathbf{h}}_L(\theta)}{\bar{\mathbf{h}}_L^H(\theta)\mathbf{R}_{n,\mathrm{H}}^{-1}\bar{\mathbf{h}}_L(\theta)} & 0\end{bmatrix}^T, \tag{26}$$

where $\mathbf{R}_{n,\mathrm{H}}$ is the $(M-1)\times(M-1)$-dimensional noise correlation matrix only using the head-mounted microphones. The MVDR pre-processed (biased) RTF estimate is then given by

$$\boxed{H_{e,L}^{\mathrm{pp}} = \frac{\mathcal{E}\left\{Y_e\mathbf{y}^H\mathbf{w}_{\mathrm{H},L}\right\}}{\mathcal{E}\left\{\mathbf{w}_{\mathrm{H},L}^H\mathbf{y}\mathbf{y}^H\mathbf{w}_{\mathrm{H},L}\right\}} = \frac{\mathbf{e}_e^T\mathbf{R}_y\mathbf{w}_{\mathrm{H},L}}{\mathbf{w}_{\mathrm{H},L}^H\mathbf{R}_y\mathbf{w}_{\mathrm{H},L}}} \tag{27}$$

By substituting $\mathbf{w}_{\mathrm{H},R}$ in (27) it can easily be shown that $H_{e,L}^{\mathrm{pp}} = H_{e,R}^{\mathrm{pp}}\frac{\bar{A}_R}{\bar{A}_L}$ and hence, by using the pre-processed estimator in

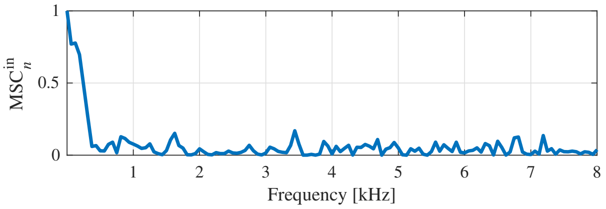

Figure 3: *Input noise MSC generated by four loudspeakers*

(27) the RTF vectors in (23) are parallel, what leads to the mapping of the residual noise to the position of the desired speech source.

4.3. Covariance whitening

Covariance whitening is a well-known approach to estimate RTFs [14, 15, 17]. The noise correlation matrix can be factorized into a lower triangular matrix $\mathbf{L}$ and its conjugate transpose $\mathbf{L}^H$ using the Cholesky decomposition, i.e., [15, 17]

$$\mathbf{R}_n = \mathbf{L}\mathbf{L}^H, \quad \mathbf{R}_n^{-1} = \mathbf{L}^{-H}\mathbf{L}^{-1}. \tag{28}$$

Using (28), the pre-whitened correlation matrix of the microphone signals is given by

$$\mathbf{R}_y^{\mathrm{w}} = \mathbf{L}^{-1}\mathbf{R}_y\mathbf{L}^{-H}. \tag{29}$$

The eigenvalue decomposition (EVD) of this pre-whitened matrix is given by

$$\mathbf{R}_y^{\mathrm{w}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H, \tag{30}$$

with $\mathbf{V}$ an $M \times M$-dimensional matrix containing the eigenvectors and $\mathbf{\Lambda}$ an $M \times M$-dimensional diagonal matrix containing the eigenvalues. Using the eigenvector $\mathbf{v}_{\max}$ that corresponds to the largest eigenvalue, the RTF can be estimated as [15, 17]

$$\boxed{H_{e,L}^{\mathrm{cw}} = \frac{\mathbf{e}_e^T\mathbf{L}\mathbf{v}_{\max}}{\mathbf{e}_L^T\mathbf{L}\mathbf{v}_{\max}}} \tag{31}$$

Compared to the MVDR pre-processed approach in Section 4.2, the covariance whitening approach aims at estimating the *reverberant* RTFs and hence, the RTF vectors in (23) are not parallel.

5. Experimental results

5.1. Setup

All signals were recorded in a laboratory with variable acoustics (7 m × 6 m × 2.7 m) where the reverberation time was set to about 350 ms. We used two behind-the-ear (BTE) hearing aid dummies each having two microphones with an inter-microphone distance of about 7.6 mm, and an external microphone, i.e., $M = 5$ microphones in total. The hearing aids were placed on the ears of a head-and-torso simulator (HATS) that was placed in the middle of the room. The desired speech source was played back by a loudspeaker placed at about 2 m distance to the middle of the head at an angle of about 35°, i.e., on the right side of the HATS. The background multi-talker noise was realized by four loudspeakers in the corners of the room that were facing the corners and playing back uncorrelated multi-talker noise. Fig. 3 shows the measured input noise MSC using the first microphone of each hearing aid as reference microphone. The speech and noise signals were recorded separately such that we were able to mix them at different input SNRs afterwards. The external microphone was placed at 0.5 m distance to the desired speech source parallel to the viewing direction of the HATS.

$\mathrm{iSNR}_R^{\mathrm{in}}$ [dB]	-10	-5	0	5
$\mathrm{iSNR}_L^{\mathrm{in}}$ [dB]	-14.5	-9.5	-4.5	0.5
$\mathrm{iSNR}_e^{\mathrm{in}}$ [dB]	-2.5	2.5	7.5	12.5

Table 1: *Input intelligibility-weighted SNRs*

For the anechoic RTF vectors $\bar{\mathbf{h}}_L(\theta)$ and $\bar{\mathbf{h}}_R(\theta)$ used in (23) we used the database presented in [24] who used similar hearing aid dummies in an anechoic room. We assumed a DOA of 35° and chose the respective measurements from the database. The processing was done at a sampling rate of 16 kHz using an STFT-based weighted overlap-add framework with a frame length of 16 ms (256 samples) and a frame shift of 50%. The input signals consisted of 2 s noise-only followed by 18 s of speech-plus-noise. The noise correlation matrix $\hat{\mathbf{R}}_n$ was estimated during the noise-only part, whereas the microphone signal correlation matrix $\hat{\mathbf{R}}_y$ was estimated during the speech-plus-noise part. The RTFs between the reference microphones and the external microphone were estimated using $\hat{\mathbf{R}}_n$ and $\hat{\mathbf{R}}_y$, cf. Section 4. The obtained filters in (22) were applied to the complete signal

We evaluated four different filters, namely

- the binaural MVDR beamformer in (26) only using the head-mounted microphones ($\mathbf{w}_{\mathrm{H}}$)
- the binaural MVDR-N beamformer in (22) using either the biased RTF estimate in (25) ($\mathbf{w}_\eta^{\mathrm{b}}$), the pre-processed RTF estimate in (27) ($\mathbf{w}_\eta^{\mathrm{pp}}$) or the covariance whitening RTF estimate in (31) ($\mathbf{w}_\eta^{\mathrm{cw}}$)

As objective performance measures we used the intelligibility-weighted SNR (iSNR) [26] improvement for the left and the right hearing aid relative to the reference microphone signals, the MSC error comparing the input noise MSC (cf. Fig. 3) with the output noise MSC, and the ILD and ITD errors comparing the input speech ILD and ITD with the output speech ILD and ITD. All measures have been averaged over all frequencies. We set up two experiments where we changed either the input iSNR or the mixing paramter η.

5.2. Experiment 1

In the first experiment we varied the input iSNR in the right reference microphone ($\mathrm{iSNR}_R^{\mathrm{in}}$) from -10 dB to 5 dB in steps of 5 dB. This led to the input iSNRs for the left reference microphone ($\mathrm{iSNR}_L^{\mathrm{in}}$) and the eMic ($\mathrm{iSNR}_e^{\mathrm{in}}$) as shown in Table 1. The mixing parameter was set to $\eta = 0$ such that the filter in (22) is equal to the binaural MVDR beamformer $\mathbf{w}_{0,L}$.

The results are depicted in Fig. 4. As can be observed, the performance of the filter $\mathbf{w}_{\mathrm{H}}$ does not depend on the input iSNR, whereas for the filters that exploit an RTF estimate between the reference microphones and the eMic the input iSNR influences the performance. The binaural MVDR beamformer using the covariance whitening RTF estimate $\mathbf{w}_\eta^{\mathrm{cw}}$ clearly outperforms all other filters in all objective measures.

It can be observed especially for the right iSNR improvement that the covariance whitening RTF estimate is less affected by a low input iSNR. For all values of the right input iSNR the covariance whitening RTF estimate leads to the highest output iSNR for both the left and the right side. Further, the filter using the pre-processed RTF estimate $\mathbf{w}_\eta^{\mathrm{pp}}$ always leads to a higher output iSNR than the filter using the biased RTF estimate $\mathrm{w}_\eta^{\mathrm{b}}$. The filter $\mathbf{w}_{\mathrm{H}}$ always leads to the lowest output iSNR.

For the MSC error of the noise component the filters using

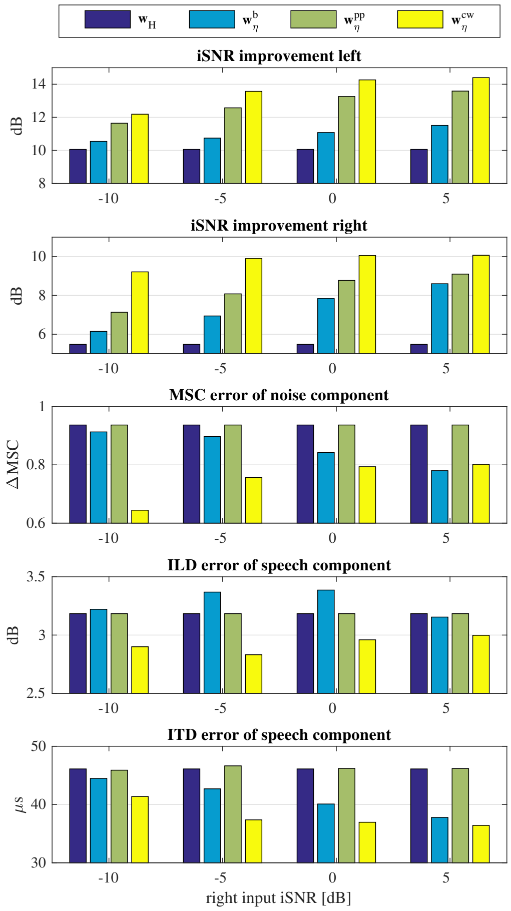

Figure 4: *Results of the first experiment where the input* iSNR *in the right reference microphone has been changed and the mixing paramter has been set to* $\eta = 0$

parallel RTF vectors ($\mathbf{w}_{\mathrm{H}}$ and $\mathbf{w}_{\eta}^{\mathrm{pp}}$) lead to a constant value, whereas the filters using non-parallel RTF vectors ($\mathbf{w}_{\eta}^{\mathrm{b}}$ and $\mathbf{w}_{\eta}^{\mathrm{cw}}$) lead to smaller errors. The MSC error of the noise component decreases with increasing right input iSNR for the filter using the biased estimate $\mathbf{w}_{\eta}^{\mathrm{b}}$ and increases for the filter using the covariance whitening estimate $\mathbf{w}_{\eta}^{\mathrm{cw}}$. While $\mathbf{w}_{\eta}^{\mathrm{cw}}$ outperforms $\mathbf{w}_{\eta}^{\mathrm{b}}$ for low right input iSNRs, the biased approach leads to the smallest MSC error for the highest right input iSNR and hence outperforms all other filters in this condition.

The ILD error of the speech component does not vary much with changes of the right input iSNR, but $\mathbf{w}_{\eta}^{\mathrm{cw}}$ outperforms all other filters in all conditions.

The ITD error of the speech component is constant over all conditions for the filters using the parallel RTF vectors ($\mathbf{w}_{\mathrm{H}}$ and $\mathbf{w}_{\eta}^{\mathrm{pp}}$) and decreasing with increasing right input iSNR for the filters using non-parallel RTF vectors ($\mathbf{w}_{\eta}^{\mathrm{b}}$ and $\mathbf{w}_{\eta}^{\mathrm{cw}}$), while $\mathbf{w}_{\eta}^{\mathrm{cw}}$ outperforms $\mathbf{w}_{\eta}^{\mathrm{b}}$.

In conclusion, it appears that even when using anechoic RTFs for the head-mounted microphones, using reverberant RTF estimates between the reference microphones and the external microphone (as in $\mathbf{w}_{\eta}^{\mathrm{b}}$ and $\mathbf{w}_{\eta}^{\mathrm{cw}}$) may lead to slight binaural cue preservation of the noise without even applying partial noise estimation.

5.3. Experiment 2

In the second experiment we set the input iSNR in the right reference microphone to -5 dB (cf. Table 1) and varied the mixing parameter η in (22) from 0 to 0.2 in steps of 0.05. The results for the second experiment are depicted in Fig. 5. The binaural MVDR beamformer using only the head-mounted microphones $\mathbf{w}_{\mathrm{H}}$ is obviously not affected by the mixing parameter η but yields a reference of the filter performance without incorporating an eMic.

In terms of iSNR improvement the performance of the binaural MVDR-N beamformer using an external microphone is better than the binaural MVDR beamformer only using the head-mounted microphones for small values of η. This effect decreases with increasing η, i.e., the output iSNR of the binaural MVDR-N beamformer is decreasing with η. For low values of η the filter using the covariance whitening RTF estimate $\mathbf{w}_{\eta}^{\mathrm{cw}}$ clearly outperforms all other filters, while for larger η the distance to the other filters decreases. Hence, it appears that η has higher influence on $\mathbf{w}_{\eta}^{\mathrm{cw}}$ than on $\mathbf{w}_{\eta}^{\mathrm{b}}$ and $\mathbf{w}_{\eta}^{\mathrm{pp}}$. The preprocessing done in $\mathbf{w}_{\eta}^{\mathrm{pp}}$ proves beneficial for all values of η compared to the filter using the biased estimate $\mathbf{w}_{\eta}^{\mathrm{b}}$.

The MSC error of the noise component is decreasing with η for the binaural MVDR-N beamformer, which is intuitively clear because more and more of the noisy reference microphone signal is added to the beamformer output. The filter $\mathbf{w}_{\eta}^{\mathrm{cw}}$ clearly outperforms all other filters, while $\mathbf{w}_{\eta}^{\mathrm{b}}$ only slightly outperforms $\mathbf{w}_{\eta}^{\mathrm{pp}}$ for very small values of η.

The ILD and ITD errors of the speech component are decreasing with increasing η for the binaural MVDR-N beamformer. Please note, that in theory the ILD and ITD errors of the speech component are equal to 0 but due to the use of anechoic RTFs these errors occur. The filter $\mathbf{w}_{\eta}^{\mathrm{cw}}$ again outperforms all other filters, while $\mathbf{w}_{\eta}^{\mathrm{pp}}$ outperforms $\mathbf{w}_{\eta}^{\mathrm{b}}$ in terms of ILD error, and $\mathbf{w}_{\eta}^{\mathrm{b}}$ outperforms $\mathbf{w}_{\eta}^{\mathrm{pp}}$ in terms of ITD error.

6. Conclusions

In this paper we investigated the influence of three different RTF estimators that estimate the RTFs between the reference microphones of two head-mounted hearing devices and an external microphone on the noise reduction and binaural cue preservation performance of the binaural MVDR-N beamformer using recorded signals. The estimator using so-called covariance whitening outperformed the other estimators. Additionaly, it appeared that using anechoic RTFs for the head-mounted microphones and reverberant RTFs for the external microphone leads to slight binaural cue preservation without even applying partial noise estimation.

7. 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*. Wiley, 2010, pp. 269–302.

[3] 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 Processing Magazine*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[4] B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, and

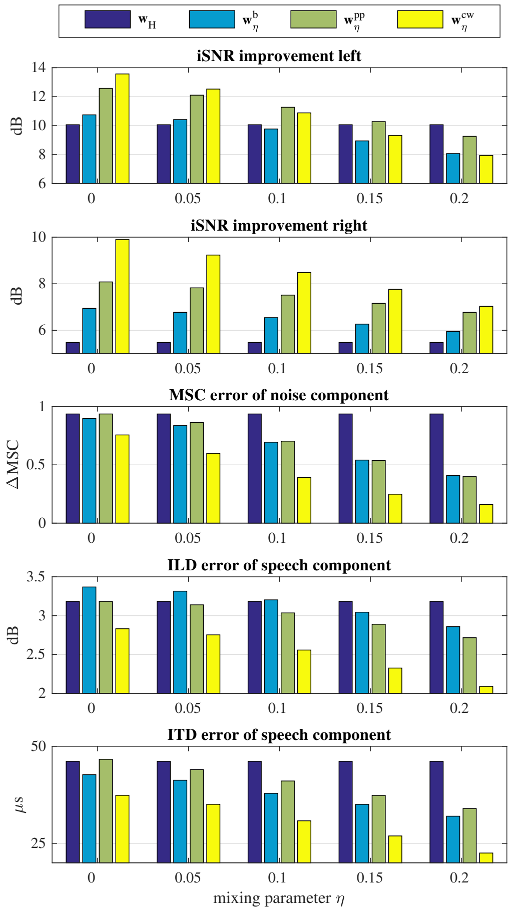

Figure 5: *Results of the second experiment where the mixing parameter η has been changed and the right input* iSNR *has been set to* -5 dB

M. Moonen, "Theoretical analysis of binaural multi-microphone noise reduction techniques," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[5] 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 Transactions on Signal Processing*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

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

[7] A. Bertrand and M. Moonen, "Robust Distributed Noise Reduction in Hearing Aids with External Acoustic Sensor Nodes," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 14 pages, Jan. 2009.

[8] N. Cvijanovic, O. Sadiq, and S. Srinivasan, "Speech enhancement using a remote wireless microphone," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 1, pp. 167–174, Feb. 2013.

[9] D. Yee, H. Kamkar-Parsi, H. Puder, and R. Martin, "A speech enhancement system using binaural hearing aids and an external microphone," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 246–250.

[10] J. Szurley, A. Bertrand, B. Van Dijk, and M. Moonen, "Binaural noise cue preservation in a binaural noise reduction system with a remote microphone signal," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 24, no. 5, pp. 952–966, May 2016.

[11] M. Farmani, M. S. Pedersen, Z.-H. Tan, and J. Jensen, "Informed sound source localization using relative transfer functions for hearing aid applications," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 25, no. 3, pp. 611–623, Jan. 2017.

[12] N. Gößling, D. Marquardt, and S. Doclo, "Performance analysis of the extended binaural MVDR beamformer with partial noise estimation in a homogeneous noise field," in *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, San Francisco, USA, Mar. 2017, pp. 1–5.

[13] R. Ali, T. van Waterschoot, and M. Moonen, "A noise reduction strategy for hearing devices using an external microphone," in *Proc. European Signal Processing Conference (EUSIPCO)*, Kos island, Greece, Aug. 2017, (submitted).

[14] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[15] 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 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015.

[16] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 451–459, Sep. 2004.

[17] I. Kodrasi and S. Doclo, "EVD-Based Multi-Channel Dereverberation of a Moving Speaker Using Different RETF Estimation Methods," in *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, San Francisco, USA, Mar. 2017, pp. 116–120.

[18] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. Cambridge, Mass. MIT Press, 1997.

[19] C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," *Journal of the Acoustical Society of America*, vol. 116, no. 5, pp. 3075–3089, 2004.

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

[21] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural Coherence Preservation in MWF-based Binaural Noise Reduction Algorithms using Partial Noise Estimation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 654–658.

[22] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical Analysis of Binaural Transfer Function MVDR Beamformers with Interference Cue Preservation Constraints," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 23, no. 12, pp. 2449–2464, Dec. 2015.

[23] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[24] 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 Journal on Advances in Signal Processing*, vol. 2009, p. 10 pages, 2009.

[25] D. P. Jarret, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: Algorithm and application," *Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1462–1472, Sep. 2012.

[26] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.