Theoretical Analysis of Linearly Constrained Multi-Channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids

Daniel Marquardt, *Student Member, IEEE*, Elior Hadad, *Student Member, IEEE*, Sharon Gannot, *Senior Member, IEEE*, 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. For the desired speech source and the interfering sources, e.g., competing speakers, this can be achieved by preserving their relative transfer functions (RTFs). It has been shown that the binaural multi-channel Wiener filter (MWF) preserves the RTF of the desired speech source, but typically distorts the RTF of the interfering sources. To this end, in this paper we propose two extensions of the binaural MWF, i.e., the binaural MWF with RTF preservation (MWF-RTF) aiming to preserve the RTF of the interfering source and the binaural MWF with interference rejection (MWF-IR) aiming to completely suppress the interfering source. Analytical expressions for the performance of the binaural MWF, MWF-RTF and MWF-IR in terms of noise reduction, speech distortion and binaural cue preservation are derived, showing that the proposed extensions yield a better performance in terms of the signal-to-interference ratio and preservation of the binaural cues of the directional interference, while the overall noise reduction performance is degraded compared to the binaural MWF. Simulation results using binaural behind-the-ear impulse responses measured in a reverberant environment validate the derived analytical expressions for the theoretically achievable performance of the binaural MWF, MWF-RTF, and MWF-IR, showing that the performance highly depends on the position of the interfering source and the number of microphones. Furthermore, the simulation results show that the MWF-RTF yields a very similar overall noise reduction performance as the binaural MWF, while preserving the binaural cues of both the speech and the interfering source.**

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

Manuscript received May 05, 2015; revised August 05, 2015; accepted September 01, 2015. Date of publication September 17, 2015; date of current version September 30, 2015. This work was supported in part by a Grant from the German-Israeli Foundation for Scientific Research and Development, in part by a joint Lower Saxony-Israeli Project financially supported by the State of Lower Saxony, Germany, and in part by 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 Dr. Roberto Togneri.

D. Marquardt and S. Doclo are with the Department of Medical Physics and Acoustics-Signal Processing Group and the Cluster of Excellence "Hearing4All," University of Oldenburg, D-26111 Oldenburg, Germany (e-mail: daniel.marquardt@uni-oldenburg.de; simon.doclo@uni-oldenburg.de).

E. Hadad and S. Gannot are with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan 5290002, Israel (e-mail: elior.hdd@gmail.com; gannotsh@gmail.com).

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.2015.2479940

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) and the interaural time difference (ITD) of all sound sources in the acoustical scene. These binaural cues play a major role in spatial awareness, i.e., for source localization and for determining the spaciousness of auditory objects [3]. Furthermore, the binaural cues are important for speech intelligibility due to binaural unmasking [3]–[6], which occurs due to a spatial separation between the desired speech and the undesired noise components. For scenarios with one desired speech source and one directional interfering source, binaural hearing can improve the speech reception threshold (SRT) by up to 12 dB in anechoic environments and up to 4 dB in reverberant environments [7]. Hence, incorporating binaural cue preservation for all sources into binaural noise reduction algorithms is an important objective.

To achieve binaural cue preservation, two main concepts for binaural noise reduction have been developed. In the first concept, the multi-channel signals from both hearing aids are used to calculate a real-valued spectral gain, where the same gain is applied to a reference microphone signal in the left and the right hearing aid [8]–[15]. This processing strategy allows perfect preservation of the instantaneous binaural cues of both the speech and the noise component, but typically suffers from a rather limited noise reduction performance and possible single-channel noise reduction artifacts [16]. The second concept is to apply a complex-valued filter to all available microphone signals on the left and the right hearing aid, combining spatial and spectral filtering [17]–[21]. 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.

In [18] the binaural multi-channel Wiener filter (MWF) has been presented, where the objective is to obtain 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 [19] that in case of a single speech source the binaural MWF preserves the relative transfer function (RTF), comprising the ILD and the ITD cues, of the speech source, but typically distorts the binaural cues of the noise component since both output components exhibit the RTF of the speech source. 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. To allow for binaural cue preservation for the noise component, an extension of the binaural MWF, namely the MWF-ITF, has been proposed in [19], by adding an additional term related to the preservation of the interaural transfer function (ITF)[1]of the noise component to the binaural MWF cost function. Furthermore, in [17] the binaural MWF with partial noise estimation (MWF-N) has been proposed, corresponding to mixing the binaural MWF output signal with a scaled version of the noisy reference microphone signals in the left and the right hearing aid. Both algorithms are able to partially preserve the binaural cues of the noise component, however, while for the MWF-ITF a trade-off between the preservation of the binaural cues of the speech and the noise component exists, for the MWF-N a trade-off between the preservation of the binaural cues of the noise component and the output signal-to-noise ratio (SNR) exists [19]. In [21] the binaural linearly constrained minimum variance (BLCMV) beamformer has been presented, which aims to partially suppress an interfering source while maintaining the RTFs of both the desired speech source and the undesired interfering source.

In this paper we consider an acoustic scenario with two speakers, i.e. one desired speech source and one interfering source, in a noisy and reverberant environment. We propose two extensions of the binaural MWF, which in addition to minimizing the overall noise output power and speech distortion aims to either preserve the binaural cues of the interfering source or to completely suppress the interfering source. The first extension, denoted as MWF-RTF, is a modification of the algorithm proposed in [20] and aims to preserve the binaural cues of the interfering source by adding a RTF preservation constraint to the binaural MWF cost function. Instead of preserving the RTF of the interfering source, one could also aim at completely suppressing the interfering source to avoid the presence of a residual interference component with distorted binaural cues in the output signal. Hence, the second extension, denoted as MWF-IR, aims to completely suppress the interfering source by adding a interference rejection constraint to the binaural MWF cost function.

The objective of this paper is to derive analytical expressions, comparing the noise reduction performance (signal-to-noise ratio, signal-to-interference ratio and signal-to-interference-plus-noise ratio) and binaural cue preservation performance of the binaural MWF and the two proposed extensions. In order to analyze the full potential of the proposed binaural MWF-based noise reduction algorithms it should be realized that the impact of estimation errors of the required signal statistics and the RTF vectors on the performance of the algorithms is not considered in this paper.

[1]Note that for the special case of a single source the ITF is equal to the RTF, e.g. as shown in [19].

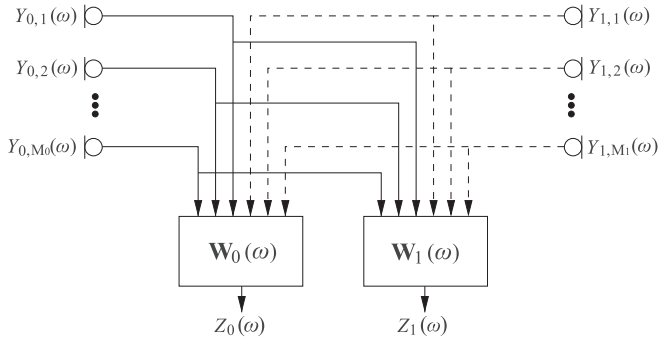


Fig. 1. General binaural hearing aid configuration, consisting of M_0 microphones on the left hearing aid and M_1 microphones on the right hearing aid. The left and the right output signals $Z_0(\omega)$ and $Z_1(\omega)$ are obtained by filtering and summing all microphone signals with the filters $\mathbf{W}_0$ and $\mathbf{W}_1$, respectively.

The paper is structured as follows. The considered signal model for a scenario with one desired speech source, one directional interfering source and additional background noise is defined in Section II. In Section III and IV, the theoretical relationship between the binaural MWF, MWF-RTF and MWF-IR will be mathematically analyzed and the performance in terms of noise reduction, speech distortion and binaural cue preservation will be thoroughly compared. In Section V, the theoretical analysis is validated by experiments using measured impulse responses of a binaural hearing aid setup in an office scenario showing that the performance of the binaural MWF, MWF-RTF and MWF-IR highly depends on the spatial position of the interfering source and the number of microphones.

II. Configuration and Notation

In this section we define the considered signal model (Section II-A) and introduce the binaural cues (Section II-B) and performance measures (Section II-C). Furthermore, we define mathematical expressions (Section II-D) which will be used in the theoretical analysis in Section III.

A. Microphone Signals and Output Signals

Consider the binaural hearing aid configuration in Fig. 1, consisting of a microphone array with $M = M_0+M_1$ microphones, with M_0 microphones on the left hearing aid and M_1 microphones on the right hearing aid. The m-th microphone signal of the left hearing aid $Y_{0,m}(\omega)$ can be written in the frequency-domain as

$$Y_{0,m}(\omega) = X_{0,m}(\omega) + U_{0,m}(\omega) + N_{0,m}(\omega), \qquad (1)$$

with $X_{0,m}(\omega)$ the speech component, $U_{0,m}(\omega)$ the interference component and $N_{0,m}(\omega)$ the background noise component in the m-th microphone signal. The m-th microphone signal of the right hearing aid $Y_{1,m}(\omega)$ is defined similarly. For conciseness

we will omit the frequency variable ω in the remainder of the paper. We define the M-dimensional stacked signal vector $\mathbf{Y}$ as

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

which can be written as

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

where the vectors $\mathbf{X}$, $\mathbf{U}$ and $\mathbf{N}$ are defined similarly as $\mathbf{Y}$ and the vector $\mathbf{V} = \mathbf{U} + \mathbf{N}$ is defined as the overall noise component, i.e. interference component plus background noise. Considering an acoustical scenario with one desired speech source S_{x} and one directional interfering source S_{i}, the components $\mathbf{X}$ and $\mathbf{U}$ can be written as

$$\mathbf{X} = S_{\mathrm{x}}\mathbf{A}, \quad \mathbf{U} = S_{\mathrm{i}}\mathbf{B}, \tag{4}$$

with $\mathbf{A}$ and $\mathbf{B}$ the acoustic transfer function (ATF) vectors between the microphones and the speech source and the interfering source, respectively. 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 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, which can be written as

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

where $\mathbf{e}_0$ and $\mathbf{e}_1$ are M-dimensional vectors with one element equal to 1 and the other elements equal to 0, i.e., $\mathbf{e}_0(1) = 1$ and $\mathbf{e}_1(M_0+1) = 1$. The reference microphone signals can then be written as

$$Y_0 = S_{\mathrm{x}}A_0 + S_{\mathrm{i}}B_0 + N_0, \tag{6}$$
$$Y_1 = S_{\mathrm{x}}A_1 + S_{\mathrm{i}}B_1 + N_1. \tag{7}$$

The correlation matrices of the speech, interference and noise components are equal to

$$\mathbf{R}_x = \mathcal{E}\left\{\mathbf{X}\mathbf{X}^H\right\} = P_s\mathbf{A}\mathbf{A}^H, \tag{8}$$
$$\mathbf{R}_u = \mathcal{E}\left\{\mathbf{U}\mathbf{U}^H\right\} = P_i\mathbf{B}\mathbf{B}^H, \tag{9}$$
$$\mathbf{R}_n = \mathcal{E}\left\{\mathbf{N}\mathbf{N}^H\right\}, \tag{10}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator, $P_s = \mathcal{E}\{|S_{\mathrm{x}}|^2\}$ and $P_i = \mathcal{E}\{|S_{\mathrm{i}}|^2\}$ denote the power spectral density (PSD) of the speech source and the interfering source, respectively. Assuming statistical independence between the components in (1), the correlation matrix of the microphone signals $\mathbf{R}_y$ can be written as

$$\mathbf{R}_y = \mathbf{R}_x + \underbrace{\mathbf{R}_u + \mathbf{R}_n}_{\mathbf{R}_v}, \tag{11}$$

with $\mathbf{R}_v$ the correlation matrix of the overall noise component which is assumed to be invertible. Furthermore, we define the cross-correlation vectors of the speech component in all microphones with the speech component in the reference microphones as

$$\mathbf{r}_{x,0} = \mathcal{E}\left\{\mathbf{X}X_0^*\right\} = \mathbf{R}_x\mathbf{e}_0 = P_s\mathbf{A}A_0^*, \tag{12}$$
$$\mathbf{r}_{x,1} = \mathcal{E}\left\{\mathbf{X}X_1^*\right\} = \mathbf{R}_x\mathbf{e}_1 = P_s\mathbf{A}A_1^*. \tag{13}$$

The output signals at the left and the right hearing aid Z_0 and Z_1 are obtained by summing a filtered version of all microphone signals, i.e.,

$$Z_0 = \mathbf{W}_0^H\mathbf{Y} = \mathbf{W}_0^H\mathbf{X} + \mathbf{W}_0^H\mathbf{U} + \mathbf{W}_0^H\mathbf{N}, \tag{14}$$
$$Z_1 = \mathbf{W}_1^H\mathbf{Y} = \mathbf{W}_1^H\mathbf{X} + \mathbf{W}_1^H\mathbf{U} + \mathbf{W}_1^H\mathbf{N}, \tag{15}$$

with $\mathbf{W}_0$ and $\mathbf{W}_1$, the filter in the left an the right hearing aid, respectively. Furthermore, we define the $2M$-dimensional complex-valued stacked weight vector $\mathbf{W}$ as

$$\mathbf{W} = \begin{bmatrix}\mathbf{W}_0 \\ \mathbf{W}_1\end{bmatrix}. \tag{16}$$

B. Relative Transfer Functions and Binaural Cues

The RTF of the speech source and the interfering source between the reference microphones on the left and the right hearing aid is defined as the ratio of the ATFs, i.e.,

$$RTF_x^{in} = \frac{A_0}{A_1}, \qquad RTF_u^{in} = \frac{B_0}{B_1}. \tag{17}$$

The output RTFs of the speech source and the interfering source are defined as the ratio of the filtered components at the left and the right hearing aid, i.e.,

$$RTF_x^{out} = \frac{\mathbf{W}_0^H\mathbf{A}}{\mathbf{W}_1^H\mathbf{A}}, \tag{18}$$
$$RTF_u^{out} = \frac{\mathbf{W}_0^H\mathbf{B}}{\mathbf{W}_1^H\mathbf{B}}. \tag{19}$$

The binaural ILD and ITD cues can be calculated from the RTF as

$$ILD = 10\log_{10}|RTF|^2, \quad ITD = \frac{\angle RTF}{\omega}, \tag{20}$$

with $\angle$ denoting the phase.

As will be shown in Section III, for the proposed algorithms we require the RTF vectors of the speech source and the interfering source, which are defined as the ATF vectors $\mathbf{A}$ and $\mathbf{B}$ normalized with the ATFs of the reference microphones, i.e.,

$$\overline{\mathbf{A}}_0 = \frac{\mathbf{A}}{A_0}, \qquad \overline{\mathbf{A}}_1 = \frac{\mathbf{A}}{A_1}, \tag{21}$$
$$\overline{\mathbf{B}}_0 = \frac{\mathbf{B}}{B_0}, \qquad \overline{\mathbf{B}}_1 = \frac{\mathbf{B}}{B_1}. \tag{22}$$

While estimating the ATF vectors $\mathbf{A}$ and $\mathbf{B}$ is known to be quite difficult [22], several methods for estimating the RTF vectors have been proposed and applied in multi-channel noise reduction algorithms, e.g. by exploiting the nonstationarity of speech signals and using generalized eigenvalue decomposition [23]–[28]. However, it should be noted that in this paper we assume the RTF vectors of the speech source and the interfering source to be known, not taking into account the impact of RTF

estimation errors when validating the derived analytical expressions in Section V.

C. Performance Measures

The *output PSD of the speech component* in the left and the right hearing aid is defined as

$$PSD_{x,0}^{\text{out}} = \mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 = P_s \left| \mathbf{W}_0^H \mathbf{A} \right|^2, \tag{23}$$

$$PSD_{x,1}^{\text{out}} = \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1 = P_s \left| \mathbf{W}_1^H \mathbf{A} \right|^2, \tag{24}$$

$$PSD_{x}^{\text{out}} = PSD_{x,0}^{\text{out}} + PSD_{x,1}^{\text{out}}. \tag{25}$$

The *output PSD of the interference component* in the left and the right hearing aid is defined as

$$PSD_{u,0}^{\text{out}} = \mathbf{W}_0^H \mathbf{R}_u \mathbf{W}_0 = P_i \left| \mathbf{W}_0^H \mathbf{B} \right|^2, \tag{26}$$

$$PSD_{u,1}^{\text{out}} = \mathbf{W}_1^H \mathbf{R}_u \mathbf{W}_1 = P_i \left| \mathbf{W}_1^H \mathbf{B} \right|^2, \tag{27}$$

$$PSD_{u}^{\text{out}} = PSD_{u,0}^{\text{out}} + PSD_{u,1}^{\text{out}}. \tag{28}$$

The *output PSD of the overall noise component* in the left and the right hearing aid is defined as

$$PSD_{v,0}^{\text{out}} = \mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0, \tag{29}$$

$$PSD_{v,1}^{\text{out}} = \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1, \tag{30}$$

$$PSD_{v}^{\text{out}} = PSD_{v,0}^{\text{out}} + PSD_{v,1}^{\text{out}}. \tag{31}$$

The *binaural Speech Distortion* (SD) is defined as the ratio of the average input PSD of the speech component in the reference microphones and the average output PSD of the speech component, i.e.,

$$SD = \frac{P_s |A_0|^2 + P_s |A_1|^2}{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}. \tag{32}$$

The *binaural output Signal-to-Interference Ratio* (SIR) is defined as the ratio of the average output PSDs of the speech component and the interference component, i.e.,

$$SIR^{out} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_0^H \mathbf{R}_u \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_u \mathbf{W}_1}. \tag{33}$$

The *binaural output Signal-to-Interference-plus-Noise Ratio* (SINR) is defined as the ratio of the average output PSDs of the speech component and the overall noise component, i.e.,

$$SINR^{out} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_0^H \mathbf{R}_v \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_v \mathbf{W}_1}. \tag{34}$$

The *binaural output Signal-to-Noise Ratio* (SNR) is defined as the ratio of the average output PSDs of the speech component and background noise component, i.e.,

$$SNR^{out} = \frac{\mathbf{W}_0^H \mathbf{R}_x \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_x \mathbf{W}_1}{\mathbf{W}_0^H \mathbf{R}_n \mathbf{W}_0 + \mathbf{W}_1^H \mathbf{R}_n \mathbf{W}_1}. \tag{35}$$

D. Mathematical Definitions

In this Section we define mathematical expressions which will be used throughout the theoretical analysis in Section III and the performance comparison in Section IV.

The speech-distortion-weighted correlation matrix $\widetilde{\mathbf{R}}_y$ is defined as

$$\widetilde{\mathbf{R}}_y = \mathbf{R}_x + \mu \mathbf{R}_v, \tag{36}$$

with μ a trade-off parameter. Applying the matrix inversion lemma to $\widetilde{\mathbf{R}}_y$ and using the rank-1 speech correlation matrix $\mathbf{R}_x$ in (8), the inverse of $\widetilde{\mathbf{R}}_y$ can be written as [18], [19]

$$\widetilde{\mathbf{R}}_y^{-1} = (\mathbf{R}_x + \mu \mathbf{R}_v)^{-1} = \frac{1}{\mu} \left[\mathbf{R}_v^{-1} - \frac{P_s \mathbf{R}_v^{-1} \mathbf{A} \mathbf{A}^H \mathbf{R}_v^{-1}}{\mu + \rho} \right], \tag{37}$$

with

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

We define the inner products of the ATF vectors of the speech and the interfering source, weighted with the inverse of the overall noise correlation matrix $\mathbf{R}_v$ as

$$\sigma_a = \mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}, \tag{39}$$

$$\sigma_{ab} = \mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{B}, \tag{40}$$

$$\sigma_b = \mathbf{B}^H \mathbf{R}_v^{-1} \mathbf{B}, \tag{41}$$

and

$$\Sigma = \frac{|\sigma_{ab}|^2}{\sigma_a \sigma_b}. \tag{42}$$

Since $\mathbf{R}_v$ is assumed to be a positive-definite Hermitian matrix, using the Cauchy-Schwartz inequality it can be shown that

$$0 \leq \Sigma \leq 1. \tag{43}$$

Furthermore, we define the inner products of the ATF vectors of the speech and the interfering source, weighted with the inverse of the speech-distortion-weighted correlation matrix $\widetilde{\mathbf{R}}_y$ as

$$\lambda_a = \mathbf{A}^H \widetilde{\mathbf{R}}_y^{-1} \mathbf{A}, \tag{44}$$

$$\lambda_{ab} = \mathbf{A}^H \widetilde{\mathbf{R}}_y^{-1} \mathbf{B}, \tag{45}$$

$$\lambda_b = \mathbf{B}^H \widetilde{\mathbf{R}}_y^{-1} \mathbf{B}, \tag{46}$$

and

$$\Gamma = \frac{|\lambda_{ab}|^2}{\lambda_a \lambda_b}. \tag{47}$$

Using (37), it can be shown that

$$\lambda_a = \frac{\sigma_a}{\mu + \rho}, \tag{48}$$

$$\lambda_{ab} = \frac{\sigma_{ab}}{\mu + \rho}, \tag{49}$$

$$\lambda_b = \frac{1}{\mu} \left(\sigma_b - \frac{P_s |\sigma_{ab}|^2}{(\mu + \rho)} \right), \tag{50}$$

and

$$\Gamma = \frac{\mu \Sigma}{\mu + \rho(1 - \Sigma)}. \tag{51}$$

Again, since $\widetilde{\mathbf{R}}_y$ is a positive-definite Hermitian matrix, using the Cauchy-Schwartz inequality it can be shown that

$$0 \leq \Gamma \leq 1. \tag{52}$$

III. Binaural Noise Reduction Techniques

In this section we review the binaural MWF and propose two extensions aiming to either preserve the RTF of the interfering source or to completely suppress the interfering source. In addition, for all algorithms analytical expressions for the output RTF, speech distortion, output SIR and output SINR are derived. It is important to note that for the sake of readability all filter expressions in this section will be derived in terms of the ATF vectors of the speech source and the interfering source but can also be written and implemented using the RTF vectors (cf. Section II-B).

A. Binaural Multi-channel Wiener Filter (MWF)

The binaural MWF [18], [19] 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. The binaural MWF cost function estimating the speech components X_0 and X_1 in the left and the right hearing aid is defined as

$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathcal{E}\left\{ \left\| \begin{bmatrix} X_0 - \mathbf{W}_0^H \mathbf{X} \\ X_1 - \mathbf{W}_1^H \mathbf{X} \end{bmatrix} \right\|^2 + \mu \left\| \begin{bmatrix} \mathbf{W}_0^H \mathbf{V} \\ \mathbf{W}_1^H \mathbf{V} \end{bmatrix} \right\|^2 \right\}, \tag{53}$$

where the parameter μ with $\mu \geq 0$ enables a trade-off between noise reduction and speech distortion. The binaural cost function in (53) can be written as

$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathbf{W}^H \mathbf{R} \mathbf{W} - \mathbf{W}^H \mathbf{r}_x - \mathbf{r}_x^H \mathbf{W} + P, \tag{54}$$

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

$$\mathbf{R} = \begin{bmatrix} \widetilde{\mathbf{R}}_y & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \widetilde{\mathbf{R}}_y \end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix} \mathbf{r}_{x,0} \\ \mathbf{r}_{x,1} \end{bmatrix}, \tag{55}$$

with $\widetilde{\mathbf{R}}_y$ defined in (36) and $\mathbf{r}_{x,0}$ and $\mathbf{r}_{x,1}$ defined in (12) and (13). The filter minimizing $J_{\mathrm{MWF}}(\mathbf{W})$ in (54) is equal to [18]

$$\mathbf{W}_{\mathrm{MWF}} = \mathbf{R}^{-1} \mathbf{r}_x \ . \tag{56}$$

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

$$\mathbf{W}_{\mathrm{MWF},0} = (\mathbf{R}_x + \mu \mathbf{R}_v)^{-1} \mathbf{r}_{x,0}, \tag{57}$$

$$\mathbf{W}_{\mathrm{MWF},1} = (\mathbf{R}_x + \mu \mathbf{R}_v)^{-1} \mathbf{r}_{x,1}. \tag{58}$$

Assuming the speech correlation matrix $\mathbf{R}_x$ to be rank-1, as already shown in [18], [19], by using (37) in (57) and (58), the binaural MWF can be decomposed into a binaural MVDR beamformer and a single-channel Wiener postfilter applied to the output of the MVDR beamformer, i.e.,

$$\mathbf{W}_{\mathrm{MWF},0} = \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_0^*, \tag{59}$$

$$\mathbf{W}_{\mathrm{MWF},1} = \frac{\rho}{\mu + \rho} \frac{\mathbf{R}_v^{-1} \mathbf{A}}{\mathbf{A}^H \mathbf{R}_v^{-1} \mathbf{A}} A_1^*, \tag{60}$$

with ρ defined in (38). Please note that the filter expressions in (59) and (60) can be rewritten in terms of the RTF vectors of the speech source $\overline{\mathbf{A}}_0$ and $\overline{\mathbf{A}}_1$ [18], [29]. The filter for the left and the right hearing aid are related by the input RTF of the speech source as

$$\mathbf{W}_{\mathrm{MWF},0} = \left(\frac{A_0}{A_1} \right)^* \mathbf{W}_{\mathrm{MWF},1}. \tag{61}$$

Substituting (61) in (18), the output RTF of the speech source is equal to the input RTF of the speech source, i.e.,

$$RTF_x^{\mathrm{out}} = \frac{A_0}{A_1} = RTF_x^{\mathrm{in}}. \tag{62}$$

Substituting (61) in (19), the output RTF of the interfering source is equal to the input RTF of the speech source, i.e.,

$$RTF_u^{\mathrm{out}} = \frac{A_0}{A_1} = RTF_x^{\mathrm{in}}. \tag{63}$$

From (62) and (63) we can conclude that the output RTF of the speech and the interfering source are the same and equal to the input RTF of the speech source, implying that both output components are perceived as directional sources coming from the speech direction, which is obviously not desired. By substituting (59) and (60) in (32), the speech distortion of the binaural MWF is equal to (cf. Appendix A-A)

$$\boxed{SD_{\mathrm{MWF}} = \frac{(\mu + \rho)^2}{\rho^2}} \tag{64}$$

which is always larger than or equal to 1. Furthermore, the output SIR of the binaural MWF can be calculated by substituting (59) and (60) in (33) as (cf. Appendix A-B)

$$\boxed{SIR_{\mathrm{MWF}}^{\mathrm{out}} = \frac{P_s \sigma_a^2}{P_i |\sigma_{ab}|^2}} \tag{65}$$

with σ_a and σ_{ab} defined in (39) and (40). By substituting (59) and (60) in (34), the output SINR of the binaural MWF is equal to [18], [19]

$$\boxed{SINR_{\mathrm{MWF}}^{\mathrm{out}} = \rho} \tag{66}$$

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

In order to control the binaural cues of the overall noise component, it has been proposed in [20] to add a linear constraint to the binaural MWF cost function, aiming to preserve the instantaneous ITF of the overall noise component. However, since for the filter in [20] an accurate estimate of the noise component is required, in this paper we propose a modified version by adding a linear constraint to the binaural MWF cost function, aiming to preserve the RTF of the interfering source, i.e.,

$$\min_{\mathbf{W}} J_{\mathrm{MWF}}(\mathbf{W}) \quad \text{subject to} \quad \frac{\mathbf{W}_0^H \mathbf{B}}{\mathbf{W}_1^H \mathbf{B}} = \frac{B_0}{B_1}. \tag{67}$$

The constraint in (67) can be written as

$$\mathbf{W}^H \mathbf{C} = 0, \tag{68}$$

with

$$\mathbf{C} = \begin{bmatrix} \mathbf{B} \\ \alpha\mathbf{B} \end{bmatrix}, \qquad \alpha = -\frac{B_0}{B_1} = -RTF_u^{\text{in}}. \tag{69}$$

Using the method of Lagrange multipliers, the solution of the constrained optimization problem in (67) is equal to [20]

$$\mathbf{W}_{\text{MWF-RTF}} = \mathbf{R}^{-1}\mathbf{r}_x - \frac{\mathbf{R}^{-1}\mathbf{C}\mathbf{C}^H\mathbf{R}^{-1}\mathbf{r}_x}{\mathbf{C}^H\mathbf{R}^{-1}\mathbf{C}}. \tag{70}$$

The stacked filter vector in (70) can further be written as (cf. Appendix B-A)

$$\mathbf{W}_{\text{MWF-RTF},0} = \mathbf{W}_{\text{MWF},0} - \kappa\tilde{\mathbf{R}}_y^{-1}\mathbf{B}, \tag{71}$$

$$\mathbf{W}_{\text{MWF-RTF},1} = \mathbf{W}_{\text{MWF},1} - \alpha\kappa\tilde{\mathbf{R}}_y^{-1}\mathbf{B}, \tag{72}$$

with

$$\kappa = \frac{P_s(A_0 + \alpha A_1)^*\sigma_a}{(1+|\alpha|^2)\,\sigma_{ab}}\Gamma. \tag{73}$$

Although not directly visible, please note that the filter expressions in (71) and (72) can be rewritten in terms of the RTF vectors of the speech source and the interfering source, i.e. $\overline{\mathbf{A}}_0$, $\overline{\mathbf{A}}_1$, $\overline{\mathbf{B}}_0$ and $\overline{\mathbf{B}}_1$. Substituting (71) and (72) in (18), the output RTF of the speech source is equal to (cf. Appendix B-B)

$$RTF_x^{\text{out}} = \frac{A_0}{A_1}\frac{1 - \frac{\Gamma}{A_0}\left(\frac{(A_0+\alpha A_1)}{1+|\alpha|^2}\right)}{1 - \frac{\alpha^*\Gamma}{A_1}\left(\frac{(A_0+\alpha A_1)}{1+|\alpha|^2}\right)}. \tag{74}$$

Hence, contrary to the binaural MWF the output RTF of the speech source of the MWF-RTF is not always preserved. Nevertheless, for the special case of $\mu = 0$ and hence $\Gamma = 0$, the output RTF of the speech source is equal to the input RTF of the speech source.

Due to the RTF constraint in (67), the output RTF of the interfering source is preserved, i.e.,

$$RTF_u^{\text{out}} = \frac{B_0}{B_1} = RTF_u^{\text{in}}. \tag{75}$$

Substituting (71) and (72) in (32), the speech distortion of the MWF-RTF is equal to (cf. Appendix B-B)

$$\boxed{SD_{\text{MWF-RTF}} = \frac{(\mu+\rho)^2}{\rho^2}\frac{1}{(1+\Gamma^2K-2\Gamma K)}} \tag{76}$$

with

$$K = \frac{|A_0 + \alpha A_1|^2}{(1+|\alpha|^2)(|A_0|^2 + |A_1|^2)}. \tag{77}$$

Realizing that the expression in (77) can be written as the square of the normalized inner product of the two vectors $\mathbf{u}$ and $\mathbf{v}$, i.e.,

$$K = \frac{\left|\mathbf{u}^H\mathbf{v}\right|^2}{\|\mathbf{u}\|^2\|\mathbf{v}\|^2}, \tag{78}$$

with

$$\mathbf{u} = \begin{bmatrix} 1 \\ \alpha^* \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} A_0 \\ A_1 \end{bmatrix}, \tag{79}$$

it can be shown, by using the Cauchy-Schwarz inequality, that

$$0 \le K \le 1. \tag{80}$$

Furthermore, (76) implies that the speech distortion of the MWF-RTF is equal to the speech distortion of the binaural MWF in (64) multiplied with an additional term that depends on Γ and K.

The SIR of the MWF-RTF can be calculated by substituting (71) and (72) in (33), i.e. (cf. Appendix B-C),

$$\boxed{SIR_{\text{MWF-RTF}}^{\text{out}} = \frac{P_s\sigma_a^2}{P_i|\sigma_{ab}|^2}\frac{(1+\Gamma^2K-2\Gamma K)}{1-K}} \tag{81}$$

Similarly as for the speech distortion, the output SIR of the MWF-RTF is equal to the output SIR of the binaural MWF in (65) multiplied with an additional term that depends on Γ and K.

By substituting (71) and (72) in (34), the SINR of the MWF-RTF is equal to (cf. Appendix B-D)

$$\boxed{SINR_{\text{MWF-RTF}}^{\text{out}} = \rho\frac{1+\Gamma^2K-2\Gamma K}{1+\nu\Gamma^2K-2\Gamma K}} \tag{82}$$

with

$$\nu = \frac{(\mu+\rho)^2}{\mu^2\Sigma} - \frac{\rho^2+2\mu\rho}{\mu^2}. \tag{83}$$

Again, similarly as for the speech distortion and the output SIR, the output SINR of the MWF-RTF is equal to the output SINR of the binaural MWF in (66) multiplied with an additional term that depends on Γ, K and ν.

C. Binaural MWF with Interference Rejection (MWF-IR)

Instead of preserving the RTF of the interfering source as proposed in Section III-B, one could also aim at completely suppressing the interfering source in order to avoid the presence of a residual interference component with distorted binaural cues in the output signal. Hence, similarly to the BLCMV beamformer in [21] we propose to extend the binaural MWF cost function with an interference rejection constraint. The cost function for the left and the right hearing aid can be written as

$$\min_{\mathbf{W}_0} \quad J_{\text{MWF}}(\mathbf{W}_0) \quad \text{subject to} \quad \mathbf{W}_0^H\mathbf{B} = 0, \tag{84}$$

$$\min_{\mathbf{W}_1} \quad J_{\text{MWF}}(\mathbf{W}_1) \quad \text{subject to} \quad \mathbf{W}_1^H\mathbf{B} = 0, \tag{85}$$

with

$$J_{\text{MWF}}(\mathbf{W}_0) = \mathcal{E}\left\{\left\|X_0 - \mathbf{W}_0^H\mathbf{X}\right\|^2 + \mu\left\|\mathbf{W}_0^H\mathbf{V}\right\|^2\right\}, \tag{86}$$

$$J_{\text{MWF}}(\mathbf{W}_1) = \mathcal{E}\left\{\left\|X_1 - \mathbf{W}_1^H\mathbf{X}\right\|^2 + \mu\left\|\mathbf{W}_1^H\mathbf{V}\right\|^2\right\}. \tag{87}$$

The linear constraints in (84) and (85) can again be written as (68), with

$$\mathbf{C} = \mathbf{B}. \tag{88}$$

Hence, the solution to the optimization problem in (84) and (85) can be obtained from (70) by replacing $\mathbf{R}$ with $\widetilde{\mathbf{R}}_y$, $\mathbf{C}$ with $\mathbf{B}$ and $\mathbf{r}_x$ with $\mathbf{r}_{x,0}$ and $\mathbf{r}_{x,1}$, respectively, i.e.,

$$\mathbf{W}_{\mathrm{MWF-IR},0} = \tilde{\mathbf{R}}_y^{-1}\mathbf{r}_{x,0} - \frac{\tilde{\mathbf{R}}_y^{-1}\mathbf{B}\mathbf{B}^H\tilde{\mathbf{R}}_y^{-1}\mathbf{r}_{x,0}}{\mathbf{B}^H\tilde{\mathbf{R}}_y^{-1}\mathbf{B}}, \quad (89)$$

$$\mathbf{W}_{\mathrm{MWF-IR},1} = \tilde{\mathbf{R}}_y^{-1}\mathbf{r}_{x,1} - \frac{\tilde{\mathbf{R}}_y^{-1}\mathbf{B}\mathbf{B}^H\tilde{\mathbf{R}}_y^{-1}\mathbf{r}_{x,1}}{\mathbf{B}^H\tilde{\mathbf{R}}_y^{-1}\mathbf{B}}. \quad (90)$$

Using (12), (13), (45), (46), (57), (58) and (132), the expressions in (89) and (90) can further be written as

$$\mathbf{W}_{\mathrm{MWF-IR},0} = \mathbf{W}_{\mathrm{MWF},0} - \gamma A_0^*\tilde{\mathbf{R}}_y^{-1}\mathbf{B}, \quad (91)$$

$$\mathbf{W}_{\mathrm{MWF-IR},1} = \mathbf{W}_{\mathrm{MWF},1} - \gamma A_1^*\tilde{\mathbf{R}}_y^{-1}\mathbf{B}, \quad (92)$$

with

$$\gamma = \frac{P_s\sigma_a}{\sigma_{ab}}\Gamma. \quad (93)$$

Please note that the filter expressions in (91) and (92) can be rewritten in terms of the RTF vectors of the speech source and the interfering source, i.e. $\overline{\mathbf{A}}_0$, $\overline{\mathbf{A}}_1$, $\overline{\mathbf{B}}_0$ and $\overline{\mathbf{B}}_1$. Similarly as for the binaural MWF, the filter for the left and the right hearing aids are related by the input RTF of the speech component, i.e.,

$$\mathbf{W}_{\mathrm{MWF-IR},0} = \left(\frac{A_0}{A_1}\right)^*\mathbf{W}_{\mathrm{MWF-IR},1}. \quad (94)$$

Furthermore, note the similarity of the MWF-IR filters in (91) and (92) with the MWF-RTF filters in (71) and (72). Using (73) in (71) and (72) and using (93) in (91) and (92), it can be shown that for the special case

$$\kappa = \gamma A_0^*, \quad (95)$$

$$\alpha\kappa = \gamma A_1^*, \quad (96)$$

the MWF-RTF filter vectors in (71) and (72) are equal to the MWF-IR filter vectors in (91) and (92). By substituting (95) in (96) it can be shown that this is the case for

$$\alpha_s = \frac{A_1^*}{A_0^*}. \quad (97)$$

Furthermore, by substituting (97) in (77), it can be shown that

$$K_s = \frac{|A_0 + \frac{A_1^*}{A_0^*}A_1|^2}{(1+|\frac{A_1^*}{A_0^*}|^2)(|A_0|^2+|A_1|^2)} = 1. \quad (98)$$

Hence, by using (98), the analytical expressions for the speech distortion and output SINR for the MWF-IR can be easily obtained by setting $K = 1$ in the analytical expressions for the MWF-RTF in (76) and (82).

Using (94) in (18), the output RTF of the speech source for the MWF-IR is equal to

$$RTF_x^{\mathrm{out}} = \frac{A_0}{A_1} = RTF_x^{\mathrm{in}}. \quad (99)$$

Hence, contrary to the MWF-RTF the MWF-IR always preserves the RTF of the speech source, independent of the trade-off parameter μ. The output RTF of the interfering source can not be calculated since theoretically the interfering source is completely suppressed and hence not present in the output signal of the MWF-IR. Setting $K = 1$ in the expression of the speech distortion for the MWF-RTF in (76), the speech distortion of the MWF-IR is equal to

$$\boxed{SD_{\mathrm{MWF-IR}} = \frac{(\mu+\rho)^2}{\rho^2}\frac{1}{(1-\Gamma)^2}} \quad (100)$$

Since the interfering source is completely suppressed, the output SIR of the MWF-IR is equal to

$$\boxed{SIR_{\mathrm{MWF-IR}}^{\mathrm{out}} = \infty} \quad (101)$$

Setting $K = 1$ in the expression of the output SINR for the MWF-RTF in (82), the output SINR of the MWF-IR is equal to

$$\boxed{SINR_{\mathrm{MWF-IR}}^{\mathrm{out}} = \rho\frac{1+\Gamma^2-2\Gamma}{1+\nu\Gamma^2-2\Gamma}} \quad (102)$$

with ν defined in (83).

IV. Theoretical Performance Comparison Between the Binaural MWF, MWF-RTF and MWF-IR

In this section we compare the theoretical performance of the binaural MWF, MWF-RTF and MWF-IR in terms of speech distortion, output SIR, output SINR and output SNR using the analytical expressions derived in Section III.

A. Speech Distortion

Noting the similarity of the analytical expressions for the speech distortion of the MWF-RTF and MWF-IR in (76) and (100) and using the fact that $0 \leq \Gamma \leq 1$ (cf. (52)) and $0 \leq K \leq 1$ (cf. (80)), we can show that

$$(1-\Gamma)^2 \leq 1 + \Gamma^2 K - 2\Gamma K \leq 1. \quad (103)$$

Using (103) in (64), (76) and (100) we can show that the speech distortion of the presented algorithms is related as

$$\boxed{1 \leq SD_{\mathrm{MWF}} \leq SD_{\mathrm{MWF-RTF}} \leq SD_{\mathrm{MWF-IR}}} \quad (104)$$

Hence, all algorithms introduce a speech distortion greater than or equal to 1, where the MWF-IR introduces the highest amount of speech distortion and the binaural MWF introduces the lowest amount of speech distortion. The speech distortion introduced by the MWF-RTF lies between the speech distortion of the binaural MWF and the MWF-IR.

B. Signal-to-Interference Ratio

Noting the similarity of the analytical expressions for the SIR of the binaural MWF and the MWF-RTF in (65) and (81) and using the fact that $0 \leq K \leq 1$ (cf. (80)), we can show that

$$1 \leq \frac{1+\Gamma^2K-2\Gamma K}{1-K}. \quad (105)$$

Hence, the output SIR of the binaural MWF is always smaller than or equal to the output SIR of the MWF-RTF, i.e.,

$$SIR_{\mathrm{MWF}}^{\mathrm{out}} \leq SIR_{\mathrm{MWF-RTF}}^{\mathrm{out}}. \quad (106)$$

Since for the MWF-IR the interference component is completely suppressed and hence the output SIR is equal to infinity (cf. (101)), the output SIR of the presented algorithms is related as

$$\boxed{SIR^{\text{out}}_{\text{MWF}} \leq SIR^{\text{out}}_{\text{MWF-RTF}} \leq SIR^{\text{out}}_{\text{MWF-IR}}} \tag{107}$$

C. Signal-to-Interference-plus-Noise Ratio

Noting the similarity of the analytical expressions for the SINR of the binaural MWF and MWF-RTF in (66) and (82) we first show that ν defined in (83) is greater than or equal to 1, i.e.,

$$1 \leq \frac{(\mu+\rho)^2}{\mu^2\Sigma} - \frac{\rho^2+2\mu\rho}{\mu^2}. \tag{108}$$

The inequality to be proven in (108) can further be written as

$$\Sigma(\mu+\rho)^2 \leq (\mu+\rho)^2. \tag{109}$$

Since $0 \leq \Sigma \leq 1$ (cf. (43)) the inequality in (109) holds. Since $\nu \geq 1$ (cf. (108)) and $0 \leq K \leq 1$ (cf. (80)), we can now show that

$$\frac{1+\Gamma^2K-2\Gamma K}{1+\nu\Gamma^2K-2\Gamma K} \leq 1, \tag{110}$$

and hence

$$SINR^{\text{out}}_{\text{MWF-RTF}} \leq SINR^{\text{out}}_{\text{MWF}}. \tag{111}$$

In the last proof of this section we will show that the output SINR of the MWF-RTF in (82) is greater than or equal to the output SINR of the MWF-IR in (102) by showing that the following inequality holds:

$$\frac{1+\Gamma^2-2\Gamma}{1+\nu\Gamma^2-2\Gamma} \leq \frac{1+\Gamma^2K-2\Gamma K}{1+\nu\Gamma^2K-2\Gamma K}. \tag{112}$$

Due to the common terms in the output SINR of the MWF-RTF in (82) and the output SINR of the MWF-IR in (102), by using the substitutions

$$a = \Gamma^2-2\Gamma, \quad b = \nu\Gamma^2-2\Gamma, \tag{113}$$

the expression in (112) can be written as

$$\frac{1+a}{1+b} \leq \frac{1+Ka}{1+Kb}. \tag{114}$$

Since $0 \leq \Gamma \leq 1$ (cf. (52)) and $\nu \geq 1$ we can show that

$$a \geq 0, \quad b \geq 0, \quad a \leq b. \tag{115}$$

Using (115) and the fact that $0 \leq K \leq 1$ (cf. (80)), it can be shown that the inequality in (114) holds and hence, the output SINR of the presented algorithms is related as

$$\boxed{SINR^{\text{out}}_{\text{MWF-IR}} \leq SINR^{\text{out}}_{\text{MWF-RTF}} \leq SINR^{\text{out}}_{\text{MWF}}} \tag{116}$$

D. Signal-to-Noise Ratio

The performance comparison of the binaural MWF, MWF-RTF and MWF-IR for the output SNR can be derived from the performance comparison for the output SIR and the output SINR in Sections IV-B and Section IV-C. Using the definitions in (33), (34) and (35), it can be shown that

$$\frac{1}{SNR^{\text{out}}} = \frac{1}{SINR^{\text{out}}} - \frac{1}{SIR^{\text{out}}}. \tag{117}$$

Hence, using (107) and (116), the output SNR of the binaural MWF is always larger than or equal to the output SNR of the MWF-RTF, which itself is always larger than or equal to the output SNR of the MWF-IR, i.e.,

$$\boxed{SNR^{\text{out}}_{\text{MWF-IR}} \leq SNR^{\text{out}}_{\text{MWF-RTF}} \leq SNR^{\text{out}}_{\text{MWF}}} \tag{118}$$

Using the relations of the speech distortion, the output SIR and the output SINR in (104), (107) and (116), we can now conclude that for the speech distortion and the output SINR the binaural MWF shows the best performance compared to the MWF-RTF and the MWF-IR, while the MWF-RTF outperforms the MWF-IR. Although the RTF constraint in the MWF-RTF leads to a better suppression of the interfering source compared to the binaural MWF, the overall noise reduction performance, comprising the suppression of the interference component and the background noise, is degraded. In addition, the complete suppression of the interfering source in the MWF-IR leads to a degradation of the overall noise reduction performance compared to the binaural MWF and the MWF-RTF. Furthermore, using the relations of the output SIR and output SNR in (107) and (118) we can conclude that the more the interfering source is suppressed, the less suppression of the background noise can be achieved.

V. Simulations

To validate the analytical expressions derived in Section III and IV for the theoretically achievable performance, in this section we present simulation results using binaural behind-the-ear impulse responses (BTE-IRs) measured in a reverberant office environment [30]. In order to analyze the full potential of the derived algorithms, we assume that a perfect estimate of the correlation matrices and the RTF vectors of the speech source and the interfering source is available.

A. Simulation Setup and Algorithm Parameters

The performance of the three considered algorithms was evaluated using measured binaural BTE-IRs [30] at a sampling frequency of 16 kHz. Each hearing aid was equipped with 3 microphones with a distance of about 7 mm and was mounted on an artificial head. The BTE-IRs were measured both in an anechoic environment (angles ranging from -180° to 180° in steps of 5°, with the source at 3 m from the artificial head) and in an office environment with a reverberation time of approximately 300 ms (angles ranging from -90° to 90° in steps of 5°, with the source at 1 m from the artificial head). The ATF vectors $\mathbf{A}$ and $\mathbf{B}$ of the speech source and the interfering source were calculated from the measured BTE-IRs. The RTF vectors $\overline{\mathbf{A}}_0$, $\overline{\mathbf{A}}_1$, $\overline{\mathbf{B}}_0$ and $\overline{\mathbf{B}}_1$ were then calculated from the ATF vectors $\mathbf{A}$ and $\mathbf{B}$ (cf. Section II-B). The PSDs of the speech source and the interfering source P_s and P_i were calculated from two different speech signals (Welch method using FFT size of 512 and Hann window). For the background noise a cylindrically

isotropic noise field was assumed. The (i,j)-th element of the noise correlation matrix $\mathbf{R}_n^{i,j}$ was calculated using the ATFs of the anechoic BTE-IRs as

$$\mathbf{R}_n^{i,j} = P_n \frac{\sum_{l=1}^{L} H_i(\theta_l) H_j^*(\theta_l)}{\sqrt{\sum_{l=1}^{L} |H_i(\theta_l)|^2 \sum_{l=1}^{L} |H_j(\theta_l)|^2}}, \quad (119)$$

with $H(\theta_l)$ denoting the anechoic ATF at angle θ_l and L the total number of angles, i.e., $L = 72$. The PSD of the background noise P_n was equal to the PSD of speech-shaped noise. The global input SNR and the global input SIR, averaged over all frequencies, were both equal to 0 dB, leading to a global input SINR of -3 dB. The trade-off parameter μ was set to 1 for all algorithms.

B. Performance Measures

For the objective validation we calculate the global performance measures by averaging the logarithmic values of the speech distortion in (32), the output SIR in (33), the output SINR in (34), and the output SNR in (35) over all frequencies. In order to evaluate the binaural cue preservation performance, we calculate the ILD and ITD error, averaged over all frequencies for the speech and the interfering source, i.e. [19],

$$\Delta ILD = \frac{1}{L} \sum_{l=1}^{L} \left| ILD^{out}(\omega_l) - ILD^{in}(\omega_l) \right|, \quad (120)$$

$$\Delta ITD = \frac{1}{L} \sum_{l=l}^{L} \left| ITD^{out}(\omega_l) - ITD^{in}(\omega_l) \right|, \quad (121)$$

with ILD and ITD defined in (20), ω_l denoting the l-th frequency and L the total number of frequencies.

C. Results

In this section we evaluate the performance of the binaural MWF, MWF-RTF and MWF-IR in the office environment for different microphone configurations. The desired speech source was located at -35° and the position of the interfering source was varied between -90° and 90°. The interfering source position at -35° was not evaluated. In the first experiment, the performance for $M = 3$ microphones was evaluated for all performance measures and in the second experiment the performance for the performance measures global SD, global output SIR, global output SINR and global output SNR was evaluated for different number of microphones.

1) Performance for $M = 3$: In the first experiment we evaluate the performance of the binaural MWF, MWF-RTF and MWF-IR using $M = 3$ microphones, i.e. two microphones on the left hearing aid and one microphone on the right hearing aid. The global SD and global output SIR are depicted in Fig. 2(a)–2(b). The global output SIR of the MWF-IR is not depicted since it is equal to infinity. As shown in the theoretical analysis in Section III and IV, the binaural MWF introduces the lowest amount of speech distortion compared to the MWF-RTF and the MWF-IR. While the global SD of the MWF-RTF is only slightly higher than for the binaural MWF, the global SD of the MWF-IR is significantly larger, especially for interfering source positions close to the speech source position. The global output SIR of both the binaural MWF and the MWF-RTF increases for interfering source positions further away from the speech source position. It can be observed that the global output SIR of the MWF-RTF is significantly larger than the global output SIR of the binaural MWF, especially for interfering source positions far away from the speech source position. The global output SIR difference ranges from 1 dB for the interfering source position of -40° up to 7 dB for the interfering source position of 50°.

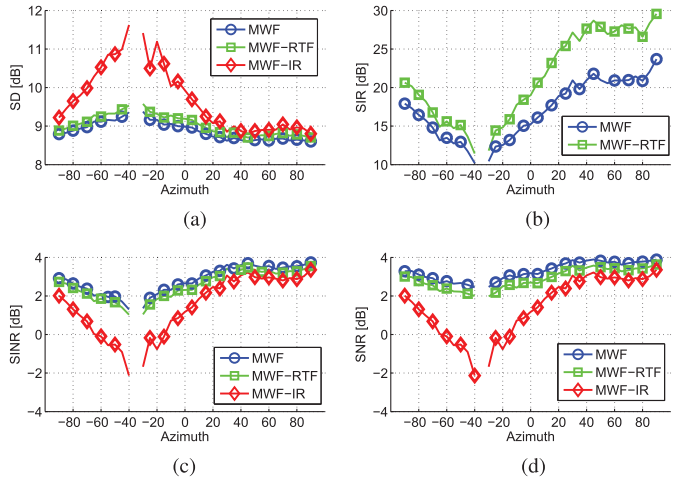


Fig. 2. Global SD, output SIR, output SINR and output SNR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at -35°, different angles of the interfering source and $M = 3$. The global input SNR and SIR are equal to 0 dB and the global input SINR is equal to -3 dB. (a) output SD (b) output SIR (c) output SINR (d) output SNR.

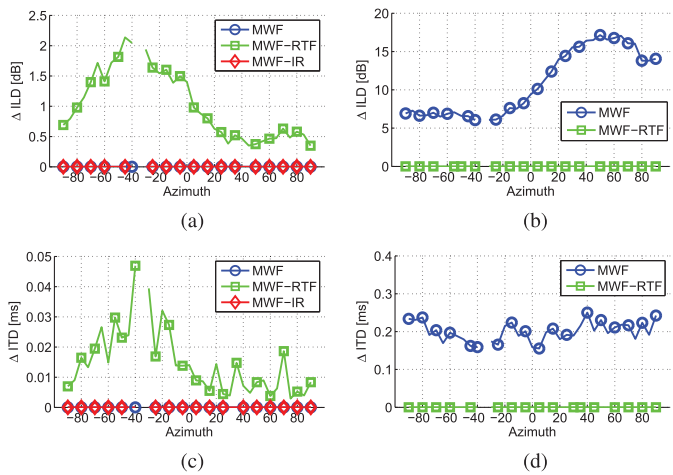


Fig. 3. Global ILD and ITD error of the speech source and the interfering source for the binaural MWF, MWF-RTF and MWF-IR for a speech source at -35°, different angles of the interfering source and $M = 3$. The global input SNR and SIR are equal to 0 dB and the global input SINR is equal to -3 dB. (a) Speech ILD error (b) Interference ILD error (c) Speech ITD error (d) Interference ITD error.

The global output SINR and global output SNR are depicted in Fig. 2(c)–2(d). The relationships between the algorithms are very similar to the results for the global SD. While the global output SINR and global output SNR for the binaural MWF and the MWF-RTF are very similar and slightly decrease for interfering source positions close to the speech source position, the global output SINR and the global output SNR for the MWF-IR is significantly lower, especially for interfering source positions close to the speech source position. The difference in global output SINR between the binaural MWF and

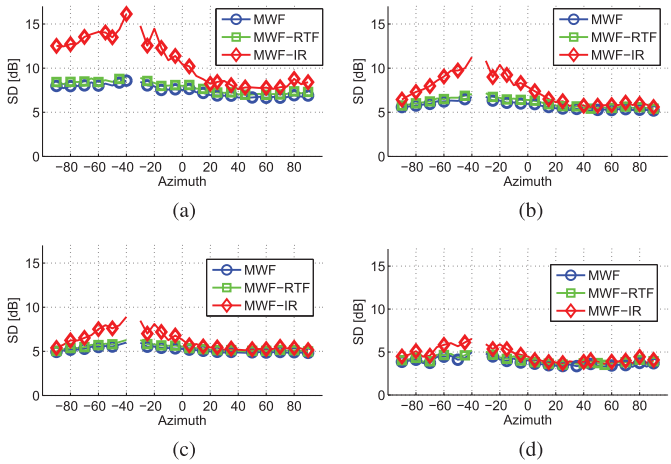


Fig. 4. Global SD for the binaural MWF, MWF-RTF and MWF-IR for a speech source at -35° and different angles of the interfering source using different number of microphones. The global input SNR and SIR are equal to 0 dB and the global input SINR is equal to -3 dB. (a) output SD, $M = 2$ (b) output SD, $M = 3$ (c) output SD, $M = 4$ (d) output SD, $M = 5$.

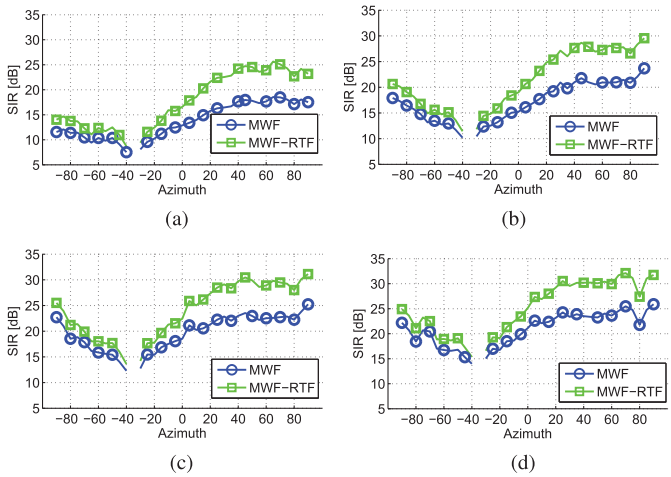


Fig. 5. Global output SIR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at -35° and different angles of the interfering source using different number of microphones. The global input SNR and SIR are equal to 0 dB and the global input SINR is equal to -3 dB. (a) output SIR, $M = 2$ (b) output SIR, $M = 3$ (c) output SIR, $M = 4$ (d) output SIR, $M = 5$.

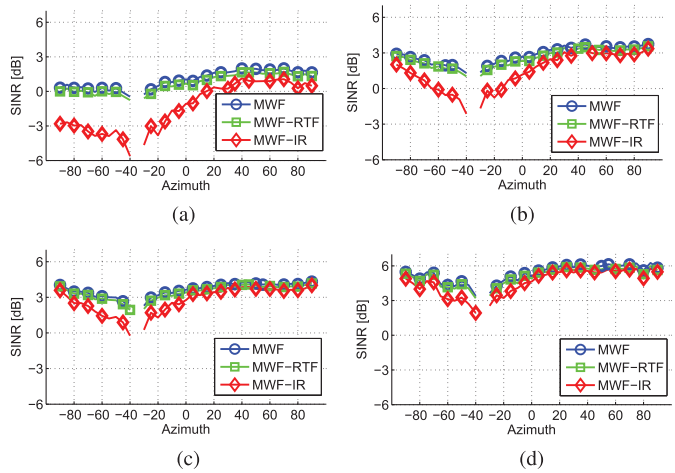


Fig. 6. Global output SINR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at -35° and different angles of the interfering source using different number of microphones. The global input SNR and SIR are equal to 0 dB and the global input SINR is equal to -3 dB. (a) output SINR, $M = 2$ (b) output SINR, $M = 3$ (c) output SINR, $M = 4$ (d) output SINR, $M = 5$.

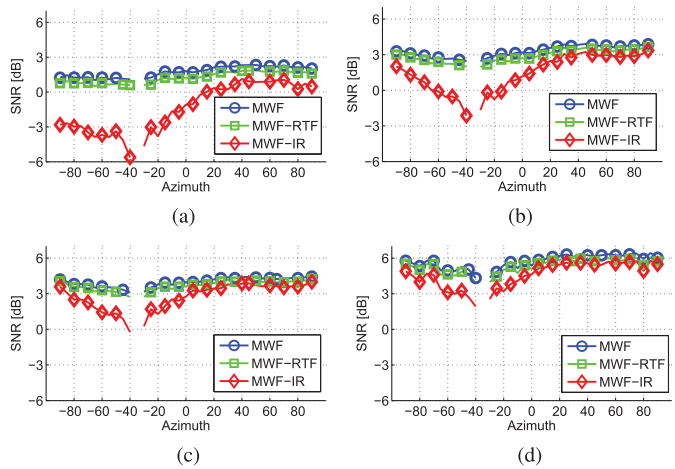


Fig. 7. Global output SNR for the binaural MWF, MWF-RTF and MWF-IR for a speech source at -35° and different angles of the interfering source using different number of microphones. The global input SNR and SIR are equal to 0 dB and the global input SINR is equal to -3 dB (a) output SNR, $M = 2$ (b) output SNR, $M = 3$ (c) output SNR, $M = 4$ (d) output SNR, $M = 5$.

the MWF-IR ranges from 0.5 dB for the interfering source position of 45° up to 3.5 dB for the interfering source position of -40°. The difference in global output SNR ranges from 0.5 dB for the interfering source position of 45° up to 4.5 dB for the interfering source position of -40°.

The ILD and ITD errors of the speech and the interference component are depicted in Fig. 3. Again, the ILD and ITD error of the MWF-IR are not depicted since the interfering source is completely suppressed. On the one hand, for the speech source the MWF-RTF introduces a small ILD error (up to 2 dB) and a very small ITD error (up to 0.05 ms), depending on the position of the interfering source, while the binaural MWF and the MWF-IR perfectly preserve the ILD and the ITD of the speech source. On the other hand, for the interfering source the binaural MWF introduces a large ILD error (up to 17 dB), especially for interfering source positions far away from the speech source position. The ITD error of the binaural MWF varies around 0.2 ms for all interfering source positions. The MWF-RTF perfectly preserves the binaural cues of the interfering source.

2) Performance for a Different Number of Microphones M : In the second experiment, we evaluate the performance of the binaural MWF, MWF-RTF and MWF-IR using $M = 2$ microphones, i.e. one microphone on the left hearing aid and one microphone on the right hearing aid, $M = 3$ microphones, i.e. two microphones on the left hearing aid and one microphone on the right hearing aid, $M = 4$ microphones, i.e. two microphones on the left hearing aid and two microphones on the right hearing aid and $M = 5$ microphones, i.e. three microphones on the left hearing aid and two microphones on the right hearing aid.

The performance measures for the binaural MWF, MWF-RTF and MWF-IR, using different number of microphones, are depicted in Fig. 4 (global SD), Fig. 5 (global output SIR), Fig. 6 (global output SINR) and Fig. 7 (global output SNR). For the binaural MWF and the MWF-RTF the global output SIR (Fig. 5) increases for an increasing number of microphones, while the performance difference between the binaural MWF and the MWF-RTF is rather independent of the number of microphones. For all algorithms and interfering

source positions, the amount of speech distortion (Fig. 4) decreases and the global output SINR (Fig. 6) and the global output SNR (Fig. 7) increase for an increasing number of microphones. Especially for $M = 2$, the performance of the MWF-IR is significantly worse than the performance of the binaural MWF and the MWF-RTF. This can be explained by the fact that for the MWF-RTF one constraint has been added to the binaural MWF cost function (cf. (67)), whereas the MWF-IR requires 2 constraints (cf. (84) and (85)), which has a severe impact on the overall performance, especially for a low number of microphones. For $M = 2$, the difference in global output SINR between the binaural MWF and the MWF-IR ranges from 1 dB for the interfering source position of 45° up to 5 dB for the interfering source position of -40°. The difference in global output SNR ranges from 1 dB for the interfering source position of 45° up to 6.5 dB for the interfering source position of -40°. It can also be observed that for an increasing number of microphones the performance of all 3 algorithms becomes more similar since the impact of the additional constraints in the MWF-RTF and the MWF-IR on the overall performance decreases.

VI. Conclusion

In this paper we have theoretically analyzed two extensions of the binaural MWF, aiming to either preserve the RTF of the interfering source (MWF-RTF) or to completely suppress the interfering source (MWF-IR). It has been shown theoretically and experimentally that for the MWF-RTF the performance in SD, output SINR and output SNR is lower but comparable to the performance of the binaural MWF, while the output SIR is larger. The MWF-IR achieves the largest SIR at the expense of an increasing speech distortion and lower output SINR and output SNR. For the MWF-RTF, the binaural cues of the interfering source are preserved, but the binaural cues of the speech source are slightly distorted, depending on the position of the interfering source. Furthermore, it has been shown that the performance for the binaural MWF and the MWF-RTF is rather independent of the position of the interfering source, whereas the performance of the MWF-IR highly depends on the position of the interfering source, especially if a small number of microphones is used. If the number of microphones is increased, the performance of the binaural MWF, MWF-RTF and MWF-IR increases and the performance difference between the algorithms becomes smaller. In future work, we will theoretically and experimentally investigate the impact of estimation errors of several RTF estimation methods on the noise reduction and binaural cue preservation performance of the proposed algorithms in different acoustic scenarios.

Appendix A
Performance of the Binaural MWF

A. Output PSD of the Speech Component and SD of the Binaural MWF

Using (59) and (60), the response of the binaural MWF to the ATF of the speech source is equal to

$$\mathbf{W}_{\mathrm{MWF},0}^{H}\mathbf{A} = \frac{A_0\rho}{\mu+\rho}, \qquad \mathbf{W}_{\mathrm{MWF},1}^{H}\mathbf{A} = \frac{A_1\rho}{\mu+\rho}. \tag{122}$$

Substituting (122) in (23) and (24), the output PSD of the speech component can be calculated as

$$PSD_x^{out} = \frac{\rho^2}{(\mu+\rho)^2} P_s \left(|A_0|^2 + |A_1|^2\right). \tag{123}$$

Substituting (123) in (32), the SD of the binaural MWF can then be calculated as

$$SD_{\mathrm{MWF}} = \frac{(\mu+\rho)^2}{\rho^2}. \tag{124}$$

B. Output PSD of the Interference Component and Output SIR of the Binaural MWF

Using (59) and (60), the response to the ATF of the interfering source is equal to

$$\mathbf{W}_{\mathrm{MWF},0}^{H}\mathbf{B} = \frac{P_s A_0 \sigma_{ab}}{\mu+\rho}, \quad \mathbf{W}_{\mathrm{MWF},1}^{H}\mathbf{B} = \frac{P_s A_1 \sigma_{ab}}{\mu+\rho}. \tag{125}$$

Substituting (125) in (26) and (27), the output PSD of the interference component can then be calculated as

$$PSD_u^{\mathrm{out}} = \frac{P_i P_s |\sigma_{ab}|^2}{(\mu+\rho)^2} P_s(|A_0|^2 + |A_1|^2). \tag{126}$$

Substituting (123) and (126) in (33), the output SIR of the binaural MWF is equal to

$$SIR_{\mathrm{MWF}}^{\mathrm{out}} = \frac{P_s \sigma_a^2}{P_i |\sigma_{ab}|^2}. \tag{127}$$

Appendix B
Performance of the MWF-RTF

A. MWF-RTF Filter Decomposition

The MWF-RTF filter in (70) is equal to

$$\mathbf{W}_{\mathrm{MWF-RTF}} = \mathbf{R}^{-1}\mathbf{r}_x - \frac{\mathbf{R}^{-1}\mathbf{C}\mathbf{C}^H\mathbf{R}^{-1}\mathbf{r}_x}{\mathbf{C}^H\mathbf{R}^{-1}\mathbf{C}}. \tag{128}$$

Using (12), (13), (45), (55) and (69) the complex-valued scalar $\mathbf{C}^H\mathbf{R}^{-1}\mathbf{r}_x$ is equal to

$$\begin{aligned}\mathbf{C}^H\mathbf{R}^{-1}\mathbf{r}_x &= \begin{bmatrix}\mathbf{B}^H\widetilde{\mathbf{R}}_y^{-1} & \alpha^*\mathbf{B}^H\widetilde{\mathbf{R}}_y^{-1}\end{bmatrix}\begin{bmatrix}P_s\mathbf{A}A_0^*\\ P_s\mathbf{A}A_1^*\end{bmatrix}, \\ &= P_s\lambda_{ab}^*(A_0+\alpha A_1)^*. \end{aligned}\tag{129}$$

Furthermore, using (46), (55) and (69) the denominator of the second term in (128) is equal to

$$\mathbf{C}^H\mathbf{R}^{-1}\mathbf{C} = \lambda_b\left(1+|\alpha|^2\right). \tag{130}$$

Hence using (56), (129) and (130) the stacked filter vector in (128) can be written as

$$\mathbf{W}_{\mathrm{MWF-RTF}} = \mathbf{W}_{\mathrm{MWF}} - \frac{P_s(A_0+\alpha A_1)^*\lambda_{ab}^*}{(1+|\alpha|^2)\,\lambda_b}\begin{bmatrix}\widetilde{\mathbf{R}}_y^{-1}\mathbf{B}\\ \alpha\widetilde{\mathbf{R}}_y^{-1}\mathbf{B}\end{bmatrix}. \tag{131}$$

By using (47), (48) and (49) it can be shown that

$$\frac{\lambda_{ab}^*}{\lambda_b} = \frac{\sigma_a}{\sigma_{ab}}\Gamma, \tag{132}$$

and hence, the MWF-RTF filter for the left and the right hearing aid are equal to

$$\mathbf{W}_{\mathrm{MWF-RTF},0} = \mathbf{W}_{\mathrm{MWF},0} - \kappa\tilde{\mathbf{R}}_y^{-1}\mathbf{B}, \quad (133)$$

$$\mathbf{W}_{\mathrm{MWF-RTF},1} = \mathbf{W}_{\mathrm{MWF},1} - \alpha\kappa\tilde{\mathbf{R}}_y^{-1}\mathbf{B}, \quad (134)$$

with

$$\kappa = \frac{P_s(A_0+\alpha A_1)^*\sigma_a}{(1+|\alpha|^2)\,\sigma_{ab}}\Gamma. \quad (135)$$

B. Output PSD of the Speech Component, Output RTF of the Speech Source and Speech Distortion of the MWF-RTF

Using (38), (49), (122), (133) and (134) the response of the MWF-RTF to the ATF of the speech source is equal to

$$\mathbf{W}^H_{\mathrm{MWF-RTF},0}\mathbf{A} = \frac{\rho}{\mu+\rho}\left(A_0 - \Gamma A_v\right), \quad (136)$$

$$\mathbf{W}^H_{\mathrm{MWF-RTF},1}\mathbf{A} = \frac{\rho}{\mu+\rho}\left(A_1 - \alpha^*\Gamma A_v\right), \quad (137)$$

with

$$A_v = \frac{(A_0+\alpha A_1)}{(1+|\alpha|^2)}. \quad (138)$$

Substituting (136) and (137) in (18), the output RTF of the speech source is equal to

$$RTF_x^{out} = \frac{A_0}{A_1}\frac{1-\Gamma\frac{A_v}{A_0}}{1-\alpha^*\Gamma\frac{A_v}{A_1}}. \quad (139)$$

Substituting (136) in (23), the output PSD of the speech component in the left hearing aid can be calculated as

$$PSD_{x,0}^{out} = \frac{P_s\rho^2}{(\mu+\rho)^2}\left[|A_0|^2 + \Gamma^2|A_v|^2 - 2\Gamma\Re\left\{A_vA_0^*\right\}\right]. \quad (140)$$

Similarly, by substituting in (137) in (24), the output PSD of the speech component in the right hearing aid can be calculated as

$$PSD_{x,1}^{\mathrm{out}} = \frac{P_s\rho^2}{(\mu+\rho)^2}\left[|A_1|^2 + \Gamma^2|\alpha|^2|A_v|^2 - 2\Gamma\Re\left\{\alpha^*A_vA_1^*\right\}\right], \quad (141)$$

and hence, using (25) the output PSD of the speech component can be written as

$$PSD_x^{out} = \frac{P_s\rho^2}{(\mu+\rho)^2}\left[|A_0|^2 + |A_1|^2 + \Gamma^2(1+|\alpha|^2)|A_v|^2 - 2\Gamma\Re\left\{(A_0+\alpha A_1)^*A_v\right\}\right]. \quad (142)$$

The expression in (142) can then further be simplified to

$$PSD_x^{\mathrm{out}} = \frac{\rho^2}{(\mu+\rho)^2}P_s\left(|A_0|^2+|A_1|^2\right)\left[1+\Gamma^2K-2\Gamma K\right], \quad (143)$$

with

$$K = \frac{|A_0+\alpha A_1|^2}{(1+|\alpha|^2)(|A_0|^2+|A_1|^2)}. \quad (144)$$

Substituting (143) in (32), the speech distortion of the MWF-RTF can then be calculated as

$$SD_{\mathrm{MWF-RTF}} = \frac{(\mu+\rho)^2}{\rho^2}\frac{1}{(1+\Gamma^2K-2\Gamma K)}. \quad (145)$$

C. Output PSD of the Interference Component and Output SIR of the MWF-RTF

Using (49), (125), (132) and (133), the response of the MWF-RTF to the ATF of the interfering source in the left hearing aid is equal to

$$\mathbf{W}^H_{\mathrm{MWF-RTF},0}\mathbf{B} = \frac{P_s\sigma_{ab}}{\mu+\rho}\alpha\left(\frac{A_0\alpha^*-A_1}{1+|\alpha|^2}\right). \quad (146)$$

Due to the RTF constraint in the MWF-RTF cost function in (67), the response to the ATF of the interfering source in the right hearing aid can be calculated as

$$\mathbf{W}^H_{\mathrm{MWF-RTF},1}\mathbf{B} = -\frac{\mathbf{W}^H_{\mathrm{MWF-RTF},0}\mathbf{B}}{\alpha} \quad (147)$$

$$= -\frac{P_s\sigma_{ab}}{\mu+\rho}\left(\frac{A_0\alpha^*-A_1}{1+|\alpha|^2}\right). \quad (148)$$

Substituting (146) and (148) in (28), the output PSD of the interference component can be calculated as

$$PSD_u^{\mathrm{out}} = \frac{P_iP_s^2|\sigma_{ab}|^2}{(\mu+\rho)^2}\frac{|A_0\alpha^*-A_1|^2}{1+|\alpha|^2}, \quad (149)$$

which, using (144), can be written as

$$PSD_u^{\mathrm{out}} = \frac{P_iP_s^2|\sigma_{ab}|^2}{(\mu+\rho)^2}(|A_0|^2+|A_1|^2)(1-K). \quad (150)$$

Substituting (143) and (150) in (33) and using (38), the output SIR of the MWF-RTF is equal to

$$SIR_{\mathrm{MWF-RTF}}^{\mathrm{out}} = \frac{P_s\sigma_a^2}{P_i|\sigma_{ab}|^2}\frac{\left(1+\Gamma^2K-2\Gamma K\right)}{1-K}. \quad (151)$$

D. Output PSD of the Overall Noise Component and Output SINR of the MWF-RTF

Using (36) and (57), the MWF-RTF filter in (133) can be written as

$$\mathbf{W}_{\mathrm{MWF-RTF},0} = \widetilde{\mathbf{R}}_y^{-1}\left(\mathbf{r}_{x,0} - \kappa\mathbf{B}\right). \quad (152)$$

Substituting (152) in (29), the output PSD of the overall noise component in the left hearing aid can be computed as

$$PSD_{v,0}^{\mathrm{out}} = \left(\mathbf{r}_{x,0}^H - \kappa^*\mathbf{B}^H\right)\mathbf{E}\left(\mathbf{r}_{x,0} - \kappa\mathbf{B}\right), \quad (153)$$

with

$$\mathbf{E} = \widetilde{\mathbf{R}}_y^{-1}\mathbf{R}_v\widetilde{\mathbf{R}}_y^{-1} \quad (154)$$

Using (37), the expression in (154) can be written as

$$\mathbf{E} = \frac{1}{\mu^2}\left[\mathbf{R}_v^{-1} - \mathbf{R}_v^{-1}\mathbf{A}\mathbf{A}^H\mathbf{R}_v^{-1}\left(\frac{P_s(\rho+2\mu)}{(\mu+\rho)^2}\right)\right]. \quad (155)$$

Using (155) in (153) and exploiting (12), (39), (40) and (41), the output PSD of the overall noise component in the left hearing aid can be written as

$$PSD_{v,0}^{\text{out}} = \frac{P_s^2|A_0|^2\sigma_a}{\mu^2}\left(1-\frac{\sigma_a P_s(\rho+2\mu)}{(\mu+\rho)^2}\right) - 2\Re\left\{\frac{P_s A_0\kappa\sigma_{ab}}{\mu^2}\left(1-\frac{\sigma_a P_s(\rho+2\mu)}{(\mu+\rho)^2}\right)\right\} + \frac{|\kappa|^2}{\mu^2}\left(\sigma_b-|\sigma_{ab}|^2\left(\frac{P_s(\rho+2\mu)}{(\mu+\rho)^2}\right)\right). \quad (156)$$

Using (38), it can be shown that

$$\frac{1}{\mu^2}\left(1-\frac{\sigma_a P_s(\rho+2\mu)}{(\mu+\rho)^2}\right) = \frac{1}{(\mu+\rho)^2}. \quad (157)$$

Hence, using (38), (42), (135) and (157), the output PSD in (156) can be written as

$$PSD_{v,0}^{\text{out}} = \frac{P_s\rho|A_0|^2}{(\mu+\rho)^2} - P_s\rho\Gamma\frac{2\Re\left\{A_0(A_0+\alpha A_1)^*\right\}}{(1+|\alpha|^2)(\mu+\rho)^2} + P_s\rho\Gamma^2\frac{|A_0+\alpha A_1|^2}{(1+|\alpha|^2)^2}\left(\frac{1}{\mu^2\Sigma}-\frac{\rho^2+2\mu\rho}{\mu^2(\mu+\rho)^2}\right). \quad (158)$$

Similarly, the output PSD of the noise component in the right hearing aid can be written as

$$PSD_{v,1}^{\text{out}} = \frac{P_s\rho|A_1|^2}{(\mu+\rho)^2} - P_s\rho\Gamma\frac{2\Re\left\{A_1\alpha(A_0+\alpha A_1)^*\right\}}{(1+|\alpha|^2)(\mu+\rho)^2} + P_s\rho\Gamma^2|\alpha|^2\frac{|A_0+\alpha A_1|^2}{(1+|\alpha|^2)^2}\left(\frac{1}{\mu^2\Sigma}-\frac{\rho^2+2\mu\rho}{\mu^2(\mu+\rho)^2}\right). \quad (159)$$

The sum of the output PSDs is then equal to

$$PSD_{v}^{\text{out}} = \frac{P_s\rho}{(\mu+\rho)^2}\left(|A_0|^2+|A_1|^2\right)\left[1+\nu\Gamma^2K-2\Gamma K\right], \quad (160)$$

with

$$\nu = \frac{(\mu+\rho)^2}{\mu^2\Sigma}-\frac{\rho^2+2\mu\rho}{\mu^2}. \quad (161)$$

Substituting (143) and (160) in (34), the output SINR of the MWF-RTF is equal to

$$SINR_{\text{MWF}-\text{RTF}}^{\text{out}} = \rho\frac{1+\Gamma^2K-2\Gamma K}{1+\nu\Gamma^2K-2\Gamma K}. \quad (162)$$

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] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

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

[11] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Robustness analysis of binaural hearing aid beamformer algorithms by means of objective perceptual quality measures," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2007, pp. 315–318.

[12] 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," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 7, pp. 1408–1419, Sep. 2009.

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

[14] 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.

[15] N. Yousefian, P. C. Loizou, and J. H. L. Hansen, "A coherence-based noise reduction algorithm for binaural hearing aids," *Speech Commun.*, vol. 58, pp. 101–110, 2014.

[16] 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: ser. Morgan & Claypool, 2013.

[17] T. Klasen, T. van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

[18] 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.

[19] 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.

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

[21] 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.

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

[23] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and non-stationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

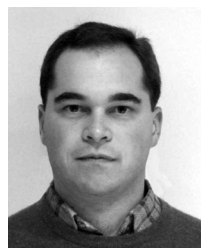
[24] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 451–459, Sep. 2004.

[25] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.
[26] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Brisbane, Australia, Apr. 2015.
[27] X. Li, L. Girin, R. Horaud, and S. Gannot, "Estimation of relative transfer function in the presence of stationary noise based on segmental power spectral density matrix subtraction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Brisbane, Australia, Apr. 2015.
[28] A. Deleforge, S. Gannot, and W. Kellermann, "Towards a generalization of relative transfer functions to more than one source," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Nice, France, Aug. 2015, accepted for publication.
[29] S. Doclo, W. Kellermann, S. Makino, and S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.
[30] 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.

Daniel Marquardt (S'12) received the Dipl.-Ing. degree in media technology with a 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.

Elior Hadad (S'13) received the B.Sc. (summa cum laude) and the M.Sc. (cum laude) degrees in electrical engineering from Ben-Gurion University of the Negev (BGU), Beer-Sheva, Israel, in 2001 and 2007, respectively. He is currently pursuing his Ph.D. degree at the Engineering Faculty, Bar-Ilan University, Israel. His research interests include statistical signal processing and in particular binaural noise reduction algorithms using microphone arrays.

Sharon Gannot (S'92–M'01–SM'06) received his B.Sc. degree (summa cum laude) from the Technion Israel Institute of Technology, Haifa, Israel, in 1986 and the M.Sc. (cum laude) and Ph.D. degrees from Tel-Aviv University, Israel in 1995 and 2000, respectively, all in electrical engineering. In 2001, he held a post-doctoral position at the department of Electrical Engineering (ESAT-SISTA) at K. U. Leuven, Belgium. From 2002 to 2003, he held a research and teaching position at the Faculty of Electrical Engineering, Technion-Israel Institute of Technology, Haifa, Israel. Currently, he is a Full Professor at the Faculty of Engineering, Bar-Ilan University, Israel, where he is heading the Speech and Signal Processing laboratory and the Signal Processing Track. Prof. Gannot is the recipient of Bar-Ilan University outstanding lecturer award for 2010 and 2014. Prof. Gannot has served as an Associate Editor of the *EURASIP Journal of Advances in Signal Processing* in 2003–2012, and as an Editor of several special issues on Multi-microphone Speech Processing of the same journal. He has also served as a guest editor of *ELSEVIER Speech Communication* and *Signal Processing* journals. Prof. Gannot has served as an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING in 2009–2013. Currently, he is a Senior Area Chair of the same journal. He also serves as a reviewer of many IEEE journals and conferences. Prof. Gannot is a member of the Audio and Acoustic Signal Processing (AASP) technical committee of the IEEE since Jan., 2010. He is also a member of the Technical and Steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) since 2005 and was the general co-chair of IWAENC held at Tel-Aviv, Israel in August 2010. Prof. Gannot has served as the general co-chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) in October 2013. Prof. Gannot was selected (with colleagues) to present a tutorial sessions in ICASSP 2012, EUSIPCO 2012, ICASSP 2013, and EUSIPCO 2013. His research interests include multi-microphone speech processing and specifically distributed algorithms for ad hoc microphone arrays for noise reduction and speaker separation; dereverberation; single microphone speech enhancement and speaker localization and tracking.



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 the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and *EURASIP Journal on Advances in Signal Processing*.