INCORPORATING RELATIVE TRANSFER FUNCTION PRESERVATION INTO THE BINAURAL MULTI-CHANNEL WIENER FILTER FOR HEARING AIDS

Daniel Marquardt[1]*, Elior Hadad*[2]*, Sharon Gannot*[2]*, Simon Doclo*[1]

[1]University of Oldenburg, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Oldenburg, Germany
[2]Bar-Ilan University, Faculty of Engineering, Ramat-Gan, Israel
daniel.marquardt@uni-oldenburg.de

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 an interfering source, e.g., competing speaker, 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 source. To this end, in this paper we propose an extension of the binaural MWF, i.e. the binaural MWF with RTF preservation (MWF-RTF) aiming to preserve the RTF of the interfering source. Analytical expressions for the performance of the binaural MWF and the MWF-RTF in terms of noise reduction and binaural cue preservation are derived, using which their performance is thoroughly compared. Simulation results using binaural behind-the-ear impulse responses measured in a reverberant environment validate the derived analytical expressions, showing that the MWF-RTF yields a better performance than the binaural MWF in terms of the signal-to-interference ratio and binaural cue preservation of the interfering source, while the overall noise reduction performance is slightly degraded.

Index Terms— Hearing aids, binaural cues, noise reduction, relative transfer function, multi-channel Wiener filter

1. 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 exploit the microphone signals from both the left and the right hearing aid are considered to be promising techniques for noise reduction, because in addition to spectral information spatial information can be exploited [1–10]. 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 confusion due to a mismatch between the acoustical and the visual information. This can be achieved by preserving the binaural cues of all sound sources in the acoustical scene.

In [5] 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 [6] that in case of a single speech source the binaural MWF preserves the Relative Transfer Function (RTF), comprising the Interaural Level Difference (ILD) and the Interaural Time Difference (ITD) cues, of the speech component. However, the binaural MWF typically distorts the binaural cues of the noise component since both output components exhibit the RTF of the speech component. Hence, after applying the binaural MWF both components are perceived as coming from the same direction and no binaural unmasking can be exploited by the auditory system. To also preserve the binaural cues of the noise component, several extensions of the binaural MWF [6, 11, 12] and the binaural LCMV beamformer [8, 13, 14] have been proposed. Since the performance of these algorithms highly depends on a careful tuning of trade-off parameters, in this paper we propose another extension of the binaural MWF, denoted as MWF-RTF, which aims to preserve the binaural cues of the interfering source by adding an RTF preservation constraint to the binaural MWF cost function. The relationship between the proposed MWF-RTF and the binaural MWF will be mathematically analysed and the performance in terms of noise reduction and binaural cue preservation will be thoroughly compared. The theoretical analysis is validated by simulation experiments using a binaural hearing aid setup in an office environment, showing that the overall noise reduction performance of the MWF-RTF is slightly degraded compared to the binaural MWF, while the signal-to-interference ratio is increased and the RTF of the interfering source is perfectly preserved.

2. CONFIGURATION AND NOTATION

Consider the binaural hearing aid configuration in Fig. 1, consisting of a microphone array with M microphones on the left and the right hearing aid. For a scenario with one desired speech source, one interfering source and background noise, 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 in the m-th micro-

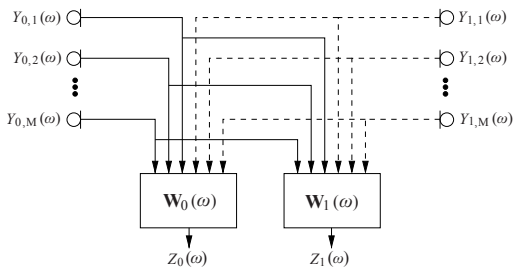


Fig. 1: Binaural hearing aid configuration

This work was supported in part by a Grant from the German-Israeli Foundation for Scientific Research and Development, a joint Lower Saxony-Israeli Project financially supported by the State of Lower Saxony, Germany and the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG).

phone 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 whenever possible in the remainder of the paper. We define the $2M$-dimensional stacked signal vector $\mathbf{Y}$ as

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

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 $\mathbf{Y}$ in (2) and the vector $\mathbf{V} = \mathbf{U} + \mathbf{N}$ denotes the overall noise component, i.e. interference component plus background noise. For the 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

$$\mathbf{X} = S_x\mathbf{A}, \qquad \mathbf{U} = S_i\mathbf{B}, \tag{4}$$

with $\mathbf{A}$ and $\mathbf{B}$ the acoustic transfer functions (ATFs) between the microphones and the speech and the interfering source, respectively. The reference microphone signals of the left and the right hearing aid 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 $2M$-dimensional vectors with one element equal to 1 and the other elements equal to 0, i.e., $\mathbf{e}_0(1) = 1$ and $\mathbf{e}_1(M+1) = 1$. The correlation matrices of the speech and the interference component are defined as

$$\mathbf{R}_x = \mathcal{E}\left\{\mathbf{X}\mathbf{X}^H\right\} = P_s\mathbf{A}\mathbf{A}^H, \;\; \mathbf{R}_u = \mathcal{E}\left\{\mathbf{U}\mathbf{U}^H\right\} = P_i\mathbf{B}\mathbf{B}^H, \tag{6}$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator, and $P_s = \mathcal{E}\left\{|S_x|^2\right\}$ and $P_i = \mathcal{E}\left\{|S_i|^2\right\}$ 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

$$\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_u + \mathbf{R}_n = \mathbf{R}_x + \mathbf{R}_v, \tag{7}$$

with $\mathbf{R}_v = \mathbf{R}_u + \mathbf{R}_n$ the correlation matrix of the overall noise component, which is assumed to be invertible, and $\mathbf{R}_n = \mathcal{E}\left\{\mathbf{N}\mathbf{N}^H\right\}$ the correlation matrix of the background noise. The output signal at the left hearing aid Z_0 is obtained by summing the filtered version of all microphone signals, i.e.,

$$Z_0 = \mathbf{W}_0^H\mathbf{Y} = \mathbf{W}_0^H\mathbf{X} + \mathbf{W}_0^H\mathbf{U} + \mathbf{W}_0^H\mathbf{N}, \tag{8}$$

with $\mathbf{W}_0$ the filter in the left hearing aid. The output signal at the right hearing aid Z_1 is similarly defined. Furthermore, we define the $4M$-dimensional complex-valued stacked filter vector $\mathbf{W}$ as

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_0 \\ \mathbf{W}_1 \end{bmatrix}. \tag{9}$$

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.,

$$RTF_x^{in} = \frac{A_0}{A_1}, \qquad RTF_u^{in} = \frac{B_0}{B_1}. \tag{10}$$

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

$$RTF_x^{out} = \frac{\mathbf{W}_0^H\mathbf{A}}{\mathbf{W}_1^H\mathbf{A}}, \qquad RTF_u^{out} = \frac{\mathbf{W}_0^H\mathbf{B}}{\mathbf{W}_1^H\mathbf{B}}. \tag{11}$$

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

$$ILD = 10\log_{10}|RTF|^2, \quad ITD = \frac{\angle RTF}{\omega}, \tag{12}$$

with $\angle$ denoting the phase.

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.,

$$SIR^{out} = \frac{\Phi_x^{\mathrm{out}}}{\Phi_u^{\mathrm{out}}} = \frac{\mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_0 + \mathbf{W}_1^H\mathbf{R}_x\mathbf{W}_1}{\mathbf{W}_0^H\mathbf{R}_u\mathbf{W}_0 + \mathbf{W}_1^H\mathbf{R}_u\mathbf{W}_1}. \tag{13}$$

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.,

$$SINR^{out} = \frac{\Phi_x^{\mathrm{out}}}{\Phi_v^{\mathrm{out}}} = \frac{\mathbf{W}_0^H\mathbf{R}_x\mathbf{W}_0 + \mathbf{W}_1^H\mathbf{R}_x\mathbf{W}_1}{\mathbf{W}_0^H\mathbf{R}_v\mathbf{W}_0 + \mathbf{W}_1^H\mathbf{R}_v\mathbf{W}_1}. \tag{14}$$

For ease of notation, we define the weighted inner products

$$\sigma_a = \mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A}, \qquad \sigma_b = \mathbf{B}^H\mathbf{R}_v^{-1}\mathbf{B}, \qquad \sigma_{ab} = \mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{B}, \tag{15}$$

and

$$\Sigma = \frac{|\sigma_{ab}|^2}{\sigma_a\sigma_b}, \qquad \text{with} \qquad 0 \le \Sigma \le 1. \tag{16}$$

3. BINAURAL NOISE REDUCTION ALGORITHMS

In this section we review the binaural MWF and propose an extension, denoted as the MWF-RTF, aiming to preserve the RTF of the interfering source. In addition, for both algorithms analytical expressions for the output RTF, the output SIR and the output SINR are derived and the theoretical performance is compared.

3.1. Binaural multi-channel Wiener filter (MWF)

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 [5, 6]

$$\boxed{J_{\mathrm{MWF}}(\mathbf{W}) = \mathcal{E}\left\{\left\|\begin{bmatrix} X_0 - \mathbf{W}_0^H\mathbf{X} \\ X_1 - \mathbf{W}_1^H\mathbf{X}\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix}\mathbf{W}_0^H\mathbf{V} \\ \mathbf{W}_1^H\mathbf{V}\end{bmatrix}\right\|^2\right\}} \tag{17}$$

where the parameter μ with $\mu \ge 0$ enables a trade-off between noise reduction and speech distortion. The binaural cost function in (17) can be written as

$$J_{\mathrm{MWF}}(\mathbf{W}) = \mathbf{W}^H\mathbf{R}\mathbf{W} - \mathbf{W}^H\mathbf{r}_x - \mathbf{r}_x^H\mathbf{W} + P, \tag{18}$$

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

$$\mathbf{R} = \begin{bmatrix} \tilde{\mathbf{R}}_y & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \tilde{\mathbf{R}}_y \end{bmatrix}, \;\; \tilde{\mathbf{R}}_y = \mathbf{R}_x + \mu\mathbf{R}_v, \;\; \mathbf{r}_x = \begin{bmatrix} \mathbf{r}_{x,0} \\ \mathbf{r}_{x,1} \end{bmatrix}, \tag{19}$$

with $\mathbf{r}_{x,0} = \mathbf{R}_x\mathbf{e}_0$ and $\mathbf{r}_{x,1} = \mathbf{R}_x\mathbf{e}_1$. The filter minimizing (18) is equal to [5]

$$\mathbf{W}_{\mathrm{MWF}} = \mathbf{R}^{-1}\mathbf{r}_x. \tag{20}$$

Applying the matrix inversion lemma to $\tilde{\mathbf{R}}_y^{-1}$ and using (6), the filter for the left and the right hearing aid can be written as [6]

$$\boxed{\mathbf{W}_{\mathrm{MWF},0} = \frac{\rho A_0^*}{\mu+\rho}\frac{\mathbf{R}_v^{-1}\mathbf{A}}{\sigma_a}, \quad \mathbf{W}_{\mathrm{MWF},1} = \frac{\rho A_1^*}{\mu+\rho}\frac{\mathbf{R}_v^{-1}\mathbf{A}}{\sigma_a}} \tag{21}$$

with σ_a defined in (15) and

$$\rho = P_s\mathbf{A}^H\mathbf{R}_v^{-1}\mathbf{A} = P_s\sigma_a. \tag{22}$$

Substituting (21) in (11), it can easily be shown that the output RTFs of both the speech source and the interfering source are equal to the input RTF of the speech source [6], i.e.,

$$RTF_x^{\mathrm{out}} = RTF_u^{\mathrm{out}} = \frac{A_0}{A_1} = RTF_x^{\mathrm{in}}, \tag{23}$$

implying that both output components are perceived as directional sources coming from the speech direction, which is obviously not desired. The output SIR of the binaural MWF can be calculated by

substituting (21) in (13), i.e.,

$$SIR_{\mathrm{MWF}}^{\mathrm{out}} = \frac{P_s \sigma_a^2}{P_i |\sigma_{ab}|^2}, \tag{24}$$

with σ_a and σ_{ab} defined in (15). Furthermore, substituting (21) in (14), the output SINR of the binaural MWF is equal to [5, 6]

$$SINR_{\mathrm{MWF}}^{\mathrm{out}} = \rho. \tag{25}$$

3.2. Binaural MWF with RTF preservation (MWF-RTF)

In order to control the binaural cues of the overall noise component, we proposed in [15] to add a constraint to the binaural MWF cost function, aiming to preserve the instantaneous interaural transfer function (ITF) of the overall noise component. However, since for the filter in [15] an accurate estimate of the overall noise component is required, in this paper we propose a modified version by adding a constraint to the binaural MWF cost function, aiming to preserve the RTF of the interfering source, i.e.,

$$\boxed{\min_{\mathbf{W}} J_{\mathrm{MWF}}(\mathbf{W}) \quad \text{subject to} \quad \frac{\mathbf{W}_0^H \mathbf{B}}{\mathbf{W}_1^H \mathbf{B}} = \frac{B_0}{B_1}} \tag{26}$$

Using (9), the constraint in (26) can be written as

$$\mathbf{W}^H \mathbf{C} = 0, \quad \mathbf{C} = \begin{bmatrix} \mathbf{B} \\ \alpha \mathbf{B} \end{bmatrix}, \qquad \alpha = -\frac{B_0}{B_1} = -RTF_u^{\mathrm{in}}. \tag{27}$$

The solution of the optimization problem in (26) is equal to [15]

$$\mathbf{W}_{\mathrm{MWF-RTF}} = \mathbf{R}^{-1}\mathbf{r}_x - \frac{\mathbf{C}^H \mathbf{R}^{-1} \mathbf{r}_x}{\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C}} \mathbf{R}^{-1} \mathbf{C}. \tag{28}$$

In addition, by defining

$$\Gamma = \frac{|\mathbf{A}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{B}|^2}{\left(\mathbf{A}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{A}\right)\left(\mathbf{B}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{B}\right)} \quad \text{and} \quad A_v = \frac{(A_0 + \alpha A_1)}{(1 + |\alpha|^2)}, \tag{29}$$

and using the matrix inversion lemma to show that

$$\mathbf{A}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{A} = \frac{\sigma_a}{\mu + \rho}, \qquad \mathbf{A}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{B} = \frac{\sigma_{ab}}{\mu + \rho}, \tag{30}$$

the stacked filter vector in (28) can be written as the sum of the binaural MWF filter vector in (21) and an additional term, i.e.,

$$\boxed{\mathbf{W}_{\mathrm{MWF-RTF},0} = \mathbf{W}_{\mathrm{MWF},0} - \kappa \tilde{\mathbf{R}}_y^{-1} \mathbf{B}} \tag{31}$$

$$\boxed{\mathbf{W}_{\mathrm{MWF-RTF},1} = \mathbf{W}_{\mathrm{MWF},1} - \alpha \kappa \tilde{\mathbf{R}}_y^{-1} \mathbf{B}} \tag{32}$$

with

$$\kappa = P_s A_v^* \frac{\mathbf{A}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{A}}{\mathbf{A}^H \tilde{\mathbf{R}}_y^{-1} \mathbf{B}} \Gamma = P_s A_v^* \frac{\sigma_a}{\sigma_{ab}} \Gamma = \frac{\rho A_v^*}{\sigma_{ab}} \Gamma. \tag{33}$$

Please note that the filter expressions in (31) and (32) can be rewritten in terms of the RTF vectors of the speech source and the interfering source, which are defined as the ATF vectors $\mathbf{A}$ and $\mathbf{B}$ normalised with the ATFs of the reference microphones. While blindly estimating the ATF vectors $\mathbf{A}$ and $\mathbf{B}$ is known to be difficult [16], several methods for blindly estimating the RTF vectors have been proposed, e.g. by exploiting the nonstationarity of speech signals [17, 18] or using the generalized eigenvalue decomposition [19–21]. However, for the sake of readability we will use the ATF formulation in the remainder of the paper.

3.3. Analytical expressions and performance comparison

In this section, we derive analytical expressions for the output RTFs of the speech and the interfering source and for the output SIR and SINR of the MWF-RTF and compare them to the analytical expressions for the binaural MWF in Section 3.1.

Using (31) and (32), the output ATF of the speech source is equal to

$$\mathbf{W}_{\mathrm{MWF-RTF},0}^H \mathbf{A} = \frac{\rho}{\mu + \rho} \left(A_0 - \Gamma A_v\right), \tag{34}$$

$$\mathbf{W}_{\mathrm{MWF-RTF},1}^H \mathbf{A} = \frac{\rho}{\mu + \rho} \left(A_1 - \alpha^* \Gamma A_v\right). \tag{35}$$

Substituting (34) and (35) in (11) and using the constraint in (26), the output RTF of the speech and the interfering source are equal to

$$RTF_x^{out} = \frac{A_0}{A_1} \frac{1 - \Gamma \frac{A_v}{A_0}}{1 - \alpha^* \Gamma \frac{A_v}{A_1}}, \quad RTF_u^{\mathrm{out}} = \frac{B_0}{B_1} = RTF_u^{\mathrm{in}}. \tag{36}$$

Hence, contrary to the binaural MWF, for the MWF-RTF the output RTF of the speech source is not perfectly preserved but the output RTF of the interfering source is perfectly preserved.

Using (34) and (35) and exploiting (6), the output PSD of the speech component Φ_x^{out} is equal to

$$\Phi_x^{\mathrm{out}} = \frac{\rho^2 P_s \left(|A_0|^2 + |A_1|^2\right)}{(\mu + \rho)^2} \left[1 + \Gamma^2 K - 2\Gamma K\right], \tag{37}$$

with

$$K = \frac{|A_0 + \alpha A_1|^2}{(1 + |\alpha|^2)(|A_0|^2 + |A_1|^2)}. \tag{38}$$

Using the Cauchy-Schwarz inequality, it can be shown that

$$0 \leq K \leq 1. \tag{39}$$

Applying similar steps for the interfering source, the output PSD of the interference component Φ_u^{out} is equal to

$$\Phi_u^{\mathrm{out}} = \frac{P_i P_s^2 |\sigma_{ab}|^2}{(\mu + \rho)^2} (|A_0|^2 + |A_1|^2)(1 - K). \tag{40}$$

Finally, combining the expressions in (37) and (40), the output SIR can be calculated, according to (13), as

$$SIR_{\mathrm{MWF-RTF}}^{\mathrm{out}} = \frac{P_s \sigma_a^2}{P_i |\sigma_{ab}|^2} \frac{\left(1 + \Gamma^2 K - 2\Gamma K\right)}{1 - K}. \tag{41}$$

Note the similarity between the analytical expressions for the SIR of the binaural MWF and the MWF-RTF in (24) and (41). Using $0 \leq K \leq 1$ (cf. (39)), we can now show that

$$1 \leq \frac{1 + \Gamma^2 K - 2\Gamma K}{1 - K}, \tag{42}$$

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

$$\boxed{SIR_{\mathrm{MWF}}^{\mathrm{out}} \leq SIR_{\mathrm{MWF-RTF}}^{\mathrm{out}}} \tag{43}$$

Using (31) and (32), the output PSD of the overall noise component in the left hearing aid for the MWF-RTF is equal to

$$\begin{aligned} \Phi_{v,0}^{\mathrm{out}} &= \mathbf{W}_{\mathrm{MWF-RTF},0}^H \mathbf{R}_v \mathbf{W}_{\mathrm{MWF-RTF},0} \\ &= \left(\mathbf{r}_{x,0}^H - \kappa^* \mathbf{B}^H\right) \mathbf{E} \left(\mathbf{r}_{x,0} - \kappa \mathbf{B}\right), \end{aligned} \tag{44}$$

with $\mathbf{E} = \tilde{\mathbf{R}}_y^{-1} \mathbf{R}_v \tilde{\mathbf{R}}_y^{-1}$. Applying the matrix inversion lemma to $\tilde{\mathbf{R}}_y^{-1}$, the matrix $\mathbf{E}$ can be written as

$$\mathbf{E} = \frac{1}{\mu^2} \left[\mathbf{R}_v^{-1} - \lambda \mathbf{R}_v^{-1} \mathbf{A} \mathbf{A}^H \mathbf{R}_v^{-1}\right] \quad \text{with} \quad \lambda = \frac{P_s(\rho + 2\mu)}{(\mu + \rho)^2}. \tag{45}$$

Using (45) in (44) and exploiting (15), the output PSD of the overall noise component in the left hearing aid can be written as

$$\begin{aligned} \Phi_{v,0}^{\mathrm{out}} = &\frac{P_s \rho |A_0|^2}{(\mu + \rho)^2} - P_s \rho \Gamma \frac{2\Re\left\{A_0 (A_0 + \alpha A_1)^*\right\}}{(1 + |\alpha|^2)(\mu + \rho)^2} + \\ &P_s \rho \Gamma^2 \frac{|A_0 + \alpha A_1|^2}{(1 + |\alpha|^2)^2} \left(\frac{1}{\mu^2 \Sigma} - \frac{\rho^2 + 2\mu\rho}{\mu^2 (\mu + \rho)^2}\right). \end{aligned} \tag{46}$$

Applying similar steps to calculate the output PSD of the noise component in the right hearing aid $\Phi_{v,1}^{\text{out}}$, the sum of the output PSDs of the overall noise component $\Phi_v^{\text{out}} = \Phi_{v,0}^{\text{out}} + \Phi_{v,1}^{\text{out}}$ is equal to

$$\Phi_v^{\text{out}} = \frac{P_s \rho}{(\mu+\rho)^2} \left(|A_0|^2 + |A_1|^2 \right) \left[1 + \nu \Gamma^2 K - 2\Gamma K \right], \quad (47)$$

with

$$\nu = \frac{(\mu+\rho)^2}{\mu^2 \Sigma} - \frac{\rho^2 + 2\mu\rho}{\mu^2}. \quad (48)$$

Finally, by substituting (37) and (47) in (14), the output SINR of the MWF-RTF is equal to

$$SINR_{\text{MWF-RTF}}^{\text{out}} = \rho \frac{1 + \Gamma^2 K - 2\Gamma K}{1 + \nu \Gamma^2 K - 2\Gamma K}. \quad (49)$$

Note the similarity between the analytical expressions for the output SINR of the binaural MWF and the MWF-RTF in (25) and (49). Using $0 \leq \Sigma \leq 1$ (cf. (16)), $0 \leq K \leq 1$ (cf. (39)) and $\mu \geq 0$, we can now show that

$$\frac{1 + \Gamma^2 K - 2\Gamma K}{1 + \nu \Gamma^2 K - 2\Gamma K} \leq 1, \quad (50)$$

such that the output SINR of the binaural MWF is always greater or equal to the output SINR of the MWF-RTF, i.e.,

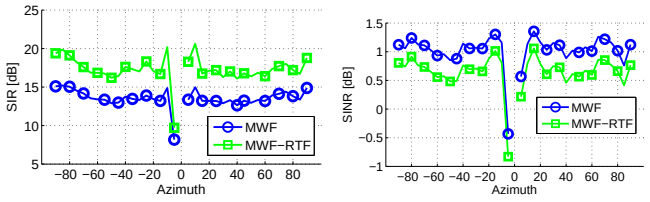
$$\boxed{SINR_{\text{MWF-RTF}}^{\text{out}} \leq SINR_{\text{MWF}}^{\text{out}}} \quad (51)$$

4. EXPERIMENTAL RESULTS

The performance of the binaural MWF and the MWF-RTF was evaluated using measured binaural behind-the-ear impulse responses (BTE-IRs) [22] at a sampling frequency of 16 kHz. Each hearing aid was equipped with 1 microphone, i.e. $M = 2$, and was mounted on an artificial head. In order to analyse 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. The BTE-IRs were measured in an office environment with a reverberation time of approximately 300 ms. The desired speech source was located at 0° and the position of the interfering source was varied between -90° and 90° in steps of 5°. The ATF vectors $\mathbf{A}$ and $\mathbf{B}$ of the speech source and the interfering source were calculated from the measured BTE-IRs. 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, where the spatial coherence matrix was calculated using anechoic ATFs of the same database [22] and the PSD of the background noise 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.

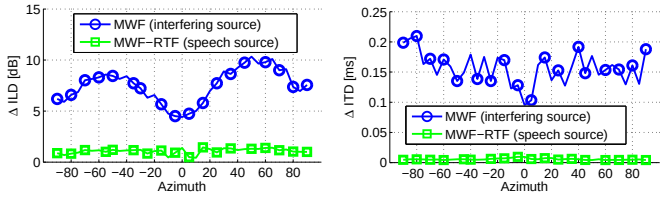
For the objective validation we have used global performance measures by averaging the logarithmic values of the output SIR in (13) and the output SINR in (14) over all frequencies. In order to evaluate the binaural cue preservation performance, we calculate the ILD and ITD error, averaged over all frequencies, for the speech and the interfering source, where the ILD and ITD cues are calculated according to (12).

The global output SIR and the global output SINR are depicted in Fig. 2. As expected from the theoretical analysis in Section 3.3, the output SIR of the MWF-RTF is always larger than the output SIR of the MWF (cf. (43)) and the output SINR of the MWF is always larger than the output SINR of the MWF-RTF (cf. (51)). Furthermore, it can be observed that the global output SIR of the MWF-RTF is significantly larger (around 5 dB) than the global output SIR of the MWF for all interfering source positions, whereas the



(a) output SIR (b) output SINR

Fig. 2: Global output SIR and SINR for the binaural MWF and the MWF-RTF for different positions of the interfering source.



(a) ILD error (b) ITD error

Fig. 3: Global ILD and ITD error for the interfering source (MWF) and for the speech source (MWF-RTF) for different positions of the interfering source.

global output SINR of the MWF is only slightly larger (around 0.5 dB) than the global output SINR of the MWF-RTF.

The ILD and ITD errors for the interfering source (MWF) and for the speech source (MWF-RTF) are depicted in Fig. 3. On the one hand, for the interfering source the MWF introduces a large ILD error (up to 10 dB) and a large ITD error (up to 0.2 ms), depending on the position of the interfering source. On the other hand, for the speech source the MWF-RTF introduces a small ILD error (up to 1.5 dB) and a very small ITD error (almost 0 ms), independent of the position of the interfering source. Please note that for the MWF the ILD and the ITD of the speech source are perfectly preserved (cf. (23)) and for the MWF-RTF the ILD and the ITD of the interfering source are perfectly preserved (cf. (36)).

In summary, from the theoretical analysis in Section 3.3 and the experimental simulation results in this section we can conclude that using the RTF constraint in the MWF-RTF leads to a significantly better suppression of the interfering source compared to the binaural MWF, while the overall noise reduction performance, comprising the suppression of both the interference component and the background noise, is slightly degraded compared to the MWF. Furthermore, while the MWF-RTF achieves perfect preservation of the binaural cues of the interfering source and only slightly distorts the binaural cues of the speech source, the MWF achieves perfect preservation of the binaural cues of the speech source but introduces a large distortion to the binaural cues of the interfering source.

5. CONCLUSION

In this paper we have theoretically analysed an extension of the binaural MWF, i.e. the MWF-RTF, which aims to preserve the RTF of the interfering source. It has been shown theoretically and experimentally that for the MWF-RTF the performance in terms of the output SINR is slightly lower but comparable to the performance of the binaural MWF, while the output SIR is significantly larger. In addition, for the MWF-RTF, the binaural cues of the interfering source are perfectly preserved, while the binaural cues of the speech source are only slightly distorted. Hence, using the MWF-RTF, preservation of the RTF of the interfering source can be incorporated into the binaural MWF without significantly degrading the overall noise reduction performance and achieving a better suppression of the interfering source.

6. REFERENCES

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

[2] T. Lotter and P. Vary, "Dual-channel speech enhancement by superdirective beamforming," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 1–14, 2006.

[3] 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 on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, Oct. 2007, pp. 315–318.

[4] K. Reindl, Y. Zheng, and W. Kellermann, "Analysis of two generic Wiener filtering concepts for binaural speech enhancement in hearing aids," in *Proc. European Signal Processing Conference (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 989–993.

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

[6] B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[7] A. H. Kamkar-Parsi and M. Bouchard, "Instantaneous Binaural Target PSD Estimation for Hearing Aid Noise Reduction in Complex Acoustic Environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1141–1154, Apr. 2011.

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

[9] M. Azarpour, G. Enzner, and R. Martin, "Binaural noise PSD estimation for binaural speech enhancement," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 2014, pp. 7068–7072.

[10] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural Coherence Preservation in Multi-channel Wiener Filtering Based Noise Reduction for Binaural Hearing Aids," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2162–2176, Dec. 2015.

[11] T.J. Klasen, T.V. 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.

[12] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Binaural multichannel Wiener filter with directional interference rejection," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, Apr. 2015, pp. 644–648.

[13] E. Hadad, S. Doclo, and S. Gannot, "The Binaural LCMV Beamformer and its Performance Analysis," *IEEE/ACM Trans. on Audio, Speech, and Lang. Proc.*, in press.

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

[15] 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 International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 21–24.

[16] Y.A. 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.

[17] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Non-Stationarity with Applications to Speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[18] 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.

[19] 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.

[20] 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.

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

[22] 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, pp. 10 pages, 2009.