Comparison of binaural multichannel Wiener filters with binaural cue preservation of the interferer

Elior Hadad¹, Daniel Marquardt², Simon Doclo² and Sharon Gannot¹
¹ Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel
² University of Oldenburg, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Oldenburg, Germany
{elior.hadad, sharon.gannot}@biu.ac.il {daniel.marquardt, simon.doclo}@uni-oldenburg.de

Abstract—An important objective of binaural speech enhancement algorithms is the preservation of the binaural cues of the sources, in addition to noise reduction. The binaural multichannel Wiener filter (MWF) preserves the binaural cues of the target but distorts the noise binaural cues. To optimally benefit from binaural unmasking and to preserve the spatial impression for the hearing aid user, two extensions of the binaural MWF have therefore been proposed, namely, the MWF with partial noise estimation (MWF-N) and MWF with interference reduction (MWF-IR). In this paper, the binaural cue preservation of these extensions is analyzed theoretically. Although both extensions are aimed at incorporating the binaural cue preservation of the interferer in the binaural MWF cost function, their properties are different. For the MWF-N, while the binaural cues of the target are preserved, there is a tradeoff between the noise reduction and the preservation of the binaural cues of the interferer component. For the MWF-IR, while the binaural cues of the interferer are preserved, those of the target may be slightly distorted. The theoretical results are validated by simulations using binaural hearing aids, demonstrating the capabilities of these beamformers in a reverberant environment.

I. INTRODUCTION

Binaural hearing devices consisting of a hearing aid mounted on each ear of a hearing-impaired person are known to outperform their monaural counterparts in terms of noise reduction performance and their capability to preserve the binaural cues and hence the spatial impression of the acoustical scene [1], [2]. By preserving the binaural cues, in addition to improving sound localization, a better speech intelligibility in noisy environments can be achieved as a result of binaural unmasking [3], [4]. For directional sources, preservation of the interaural level difference (ILD) and the interaural time difference (ITD) cues can be achieved by preserving the so-called relative transfer function (RTF), which is defined as the ratio of the acoustical transfer functions relating the source and the two ears.

In the last decade, several binaural speech enhancement algorithms aimed at preserving the binaural cues have been developed [1], [5]–[16]. In [1], the binaural MWF was presented. It was shown in [1], [8] that the binaural MWF preserves the binaural cues of the target but distorts the binaural cues of the noise, such that both the target and the noise are perceived as arriving from the target direction. Clearly, this is an undesirable phenomenon and in some scenarios (e.g., traffic) can even be dangerous to the hearing aid user.

In this paper, two extensions of the binaural MWF are discussed, which, in addition to minimizing the overall noise output power and speech distortion, are aimed at preserving the binaural cues of the interferer. The first extension, denoted as the MWF-N [8], [9], is aimed to preserve the binaural cues of the noise by mixing the output signals of the binaural MWF with a scaled version of the noisy reference microphone signals. Although in [8], [9], the aim of this beamformer was to preserve the binaural cues of the overall noise (i.e., the interferer plus the background noise), in this study, we focus on the binaural cue preservation of the directional interferer. The second extension of the binaural MWF, denoted as the MWF-IR, is aimed at preserving the binaural cues of the interferer by extending the MWF cost function with a hard constraint that controls the amount of interference reduction [13]–[15]. The binaural cue preservation of these extensions is analyzed theoretically and we compare their properties.

II. PROBLEM FORMULATION

Consider a binaural hearing aid system consisting of two hearing devices with a total of M microphones and an acoustic scenario comprising one target speech source and one directional interferer in a noisy and reverberant environment. In the frequency domain, the M-dimensional stacked vector of the received microphone signals \( y(\omega) \) can be written as

\[
y(\omega) = x(\omega) + u(\omega) + n(\omega) = x(\omega) + v(\omega),
\]

where \( x(\omega) \) is the target component, \( u(\omega) \) the interferer component, and \( n(\omega) \) the background noise component, e.g., diffuse noise. The vector \( v(\omega) = u(\omega)+n(\omega) \) is defined as the overall noise component, i.e., the interferer plus background noise component. For brevity, the frequency variable \( \omega \) is henceforth omitted.

The target and interferer components can be written as \( x = S_xa \) and \( u = S_uba \), where \( S_x \) and \( S_u \) denote the target and interferer signals and \( a \) and \( b \) denote the acoustic transfer functions (ATFs) relating the microphones and the target and the interferer, respectively.
Assuming statistical independence between all components in (1), the spatial correlation matrix of the microphone signals \( \mathbf{R}_y \) can be written as
\[
\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbf{R}_x + \mathbf{R}_u + \mathbf{R}_n = \mathbf{R}_x + \mathbf{R}_v,
\] (2)
where \( \mathbf{R}_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}, \mathbf{R}_u = \mathcal{E}\{\mathbf{u}\mathbf{u}^H\}, \) and \( \mathbf{R}_n = \mathcal{E}\{\mathbf{n}\mathbf{n}^H\} \) denote the target, interferer, and background noise correlation matrices, respectively, and \( \mathcal{E}\{\cdot\} \) is the expectation operator.

The target and interferer correlation matrices are rank-1 matrices, i.e.,
\[
\mathbf{R}_x = P_x \mathbf{a}\mathbf{a}^H, \quad \mathbf{R}_u = P_u \mathbf{b}\mathbf{b}^H,
\] (3)
where \( P_x = \mathcal{E}\{\mathbb{S}_x^2\} \) and \( P_u = \mathcal{E}\{\mathbb{S}_u^2\} \) denote the power spectral density (PSD) of the target and the interferer, respectively.

The reference microphone signals at the left and right hearing devices, selected, e.g., as the microphones closest to the ears, are given by \( y_L = e_{rL}^H \mathbf{y} \) and \( y_R = e_{rR}^H \mathbf{y} \), respectively, where \( e_{rL} \) and \( e_{rR} \) are \( M \)-dimensional indicator input vectors with one element equal to 1 and all other elements equal to zero. From (1), the reference microphone signals can be written as
\[
y_L = S_x a_L + S_n b_L + n_L, \quad y_R = S_x a_R + S_n b_R + n_R.
\] (4)
The output signals at the left and right hearing devices are given by \( z_L = w_L^H \mathbf{y} \) and \( z_R = w_R^H \mathbf{y} \), respectively, where \( w_L \) and \( w_R \) denote \( M \)-dimensional complex-valued weight filter vectors.

The RTF of the target component and the interferer component in the reference microphone of the left and the right hearing aid is defined as the ratio of the respective ATFs, i.e.,
\[
\text{RTF}^\text{an}_L = \frac{a_L}{a_R}, \quad \text{RTF}^\text{an}_R = \frac{b_L}{b_R},
\] (5)
and the output RTF is defined as the ratio of the filtered target component and the filtered interferer component in the left and the right hearing aid, respectively, i.e.,
\[
\text{RTF}^\text{out}_L = \frac{w_L^H \mathbf{a}}{w_R^H \mathbf{a}}, \quad \text{RTF}^\text{out}_R = \frac{w_L^H \mathbf{b}}{w_R^H \mathbf{b}}.
\] (6)
The binaural ILD and ITD cues can be calculated from the RTF as
\[
\text{ILD} = 10 \log_{10} |\text{RTF}|^2, \quad \text{ITD} = \frac{\angle \text{RTF}}{\omega},
\] (7)
with \( \angle \) denoting the phase.

The binaural input signal-to-interference-and-noise ratio (SINR) is defined as the ratio of the average power of the target and overall noise in the reference microphones, i.e.,
\[
\text{SINR}^\text{in} = \frac{e_{rL}^H \mathbf{R}_e e_L + e_{rR}^H \mathbf{R}_e e_R}{e_{rL}^H \mathbf{R}_e e_L + e_{rR}^H \mathbf{R}_e e_R}.
\] (8)
The binaural output SINR is defined as the ratio of the average power of the target and overall noise in the left and the right hearing aid, i.e.,
\[
\text{SINR}^\text{out} = \frac{w_L^H \mathbf{R}_e w_L + w_R^H \mathbf{R}_e w_R}{w_L^H \mathbf{R}_e w_L + w_R^H \mathbf{R}_e w_R}.
\] (9)

III. BINAURAL NOISE REDUCTION ALGORITHMS

In Section III-A, we briefly review the binaural MWF. Then, in Section III-B and in Section III-C the MWF-N and MWF-IR are described.

A. Binaural multi-channel Wiener filter (MWF)

The well-known binaural MWF produces a minimum mean square error (MSE) estimate of the target component at both reference microphones [1]. The MSE cost functions for the filter \( w_L \), estimating the target component \( x_L \) at the left hearing aid, and for the filter \( w_R \), estimating the target component \( x_R \) at the right hearing aid, are given by
\[
J_{MWF}(w_L) = \mathcal{E}\{\|x_L - w_L^H x\|^2 + \mu\|w_L^H v\|^2\},
\]
\[
J_{MWF}(w_R) = \mathcal{E}\{\|x_R - w_R^H x\|^2 + \mu\|w_R^H v\|^2\},
\] (10)
where \( \mu \) provides a weighting tradeoff between noise reduction and speech distortion. The filter vectors minimizing (10) are equal to [1]
\[
w_{L,MWF} = P_s a_L^{-1} \mathbf{R}_y, \quad w_{R,MWF} = P_s a_R^{-1} \mathbf{R}_y,
\] (11)
with \( \mathbf{R}_y = R_L + \mu \mathbf{R}_n \) defined as a speech-distortion-weighted correlation matrix. Applying the Woodbury identity to \( \mathbf{R}_y^{-1} \) and using (11), the filter vectors can be written as
\[
w_{L,MWF} = \frac{\rho}{\mu + \rho} \mathbf{R}_y^{-1} a, \quad w_{R,MWF} = \frac{\rho}{\mu + \rho} \mathbf{R}_y^{-1} a,
\] (12)
with \( \sigma_a = a^H \mathbf{R}_y^{-1} a \) and
\[
\rho = P_s a^H \mathbf{R}_y^{-1} a = P_s \sigma_a.
\] (13)
Equation (12) implies that the filter vectors of the binaural MWF \( w_{L,MWF} \) and \( w_{R,MWF} \) are parallel and the relation \( w_{L,MWF} = (\text{RTF}_x^\text{an})^* w_{R,MWF} \) holds. Hence, the RTF of the target at the output of the binaural MWF is equal to the input RTF, i.e.,
\[
\text{RTF}_x^\text{out} = \frac{w_L^H a}{w_R^H a} = a_L = \text{RTF}_x^\text{an}.
\] (14)
However, this also implies that all sound sources are perceived as arriving from the target direction, which is obviously not a desired phenomenon. Substituting (12) in (9), the output SINR of the binaural MWF is equal to [1], [8]
\[
\text{SINR}^\text{out} = \rho.
\] (15)

B. Binaural MWF with partial noise estimation (MWF-N)

An extension of the binaural MWF that, in addition to preserving the binaural cues of the target component, is also aimed to partially preserve the binaural cues of the noise component has been proposed in [9], denoted as MWF-N. The objective of the MWF-N is to produce a minimum MSE estimate of the sum of the target component and a
scaled version of the overall noise component at the reference microphones, i.e.,

\[ J_{\text{MWFS-N}}(w_L) = \mathcal{E} \{ ||x_L - w_L^H x||^2 + \mu ||\eta N v_L - w_L^H v||^2 \}, \]

\[ J_{\text{MWFS-N}}(w_R) = \mathcal{E} \{ ||x_R - w_R^H x||^2 + \mu ||\eta N v_R - w_R^H v||^2 \}, \]

where \( 0 \leq \eta N \leq 1 \) denotes the overall noise parameter, which provides a tradeoff between noise reduction and the preservation of the binaural cues of the noise component. The filter vectors minimizing (16) are equal to [8], [9]

\[ w_{L,\text{MWFS-N}} = (1 - \eta N) \frac{\rho}{\mu + \rho} \frac{R_L^{-1} a}{\sigma_a} R_L + \eta N e_L \]

\[ w_{R,\text{MWFS-N}} = (1 - \eta N) \frac{\rho}{\mu + \rho} \frac{R_R^{-1} a}{\sigma_a} R_R + \eta N e_R, \]

(17)

Hence, the output signals of the MWFS-N are equal to the sum of the output signals of the binaural MWF (weighted with \((1 - \eta N)\)) and the noisy reference microphone signals (weighted with \(\eta N\)). As shown in [8], for the MWFS-N the output RTF of the target component is equal to the input RTF for all tradeoff parameters \(\eta N\), i.e.,

\[ \text{RTF}^x_{\text{out}} = (1 - \eta N) \frac{\rho}{\mu + \rho} \frac{\rho a_L}{\rho a_L + \eta N a_L} = a_L \]

\[ \text{RTF}^u_{\text{out}} = (1 - \eta N) \frac{\rho a_R}{\rho a_R + \eta N a_R} \text{RTF}^x_{\text{in}}, \]

(18)

Substituting (17) in (6), it can be shown that the output RTF of the interferer is equal to

\[ \text{RTF}^x_{\text{out}} = (1 - \eta N) \frac{\rho a_L}{\rho a_L + \eta N a_L} \text{RTF}^x_{\text{in}}, \]

\[ + \frac{b_L}{\rho a_R + \eta N a_R} \text{RTF}^u_{\text{in}}, \]

(19)

where \( \rho a_L = P_s \sigma a_L \) and \( \sigma_{ab} = a^H b \).

Equation (19) shows that output RTF of the interferer is a weighted sum of the input RTF of the target and the input RTF of the interferer. If \( \eta N = 0 \), the output RTF of the interferer is equal to the input RTF of the target while on the other hand, if \( \eta N = 1 \), the output RTF of the interferer is equal to the input RTF of the interferer.

Substituting (17) in (9), the output SINR for the MWFS-N is equal to

\[ \text{SINR}^x_{\text{MWFS-N}} = \frac{(\mu \eta N + \rho)^2 \text{SINR}^x_{\text{in}}}{(1 - \eta N) (2 \mu \eta N + \rho + \rho \eta N) \text{SINR}^x_{\text{in}} + (\mu + \rho)^2 \eta N}, \]

(20)

where

\[ \text{SINR}^x_{\text{in}} = \frac{P_s (|a_L|^2 + |a_R|^2)}{\epsilon L R e_L + \epsilon R L e_R}. \]

(21)

1Similar derivation for the output SINRs for the left and the right hearing aids are obtained in [8].

C. Binaural MWF with interference reduction (MWF-IR)

The second extension of the binaural MWF, denoted as the MWF-IR, is aimed to better control the suppression and binaural cue preservation of the interferer by extending the MWF cost function in (10) with a hard interference reduction (IR) constraint that controls the amount of interference reduction [15], i.e.,

\[ \min_{w_L} J_{\text{MWFS}}(w_L) \text{ subject to } w_L^H b = \eta R b_L, \]

\[ \min_{w_R} J_{\text{MWFS}}(w_R) \text{ subject to } w_R^H b = \eta R b_R, \]

(22)

where \( 0 \leq \eta R \ll 1 \) denotes the interference parameter, which provides a tradeoff between interference reduction and binaural cue preservation. The filter vectors minimizing (22) are equal to

\[ w_{L,\text{MWFS-IR}} = P_s a_L \left( R_y^{-1} a - \frac{\lambda a_{ab}}{\lambda b} R_y b \right), \]

\[ w_{R,\text{MWFS-IR}} = P_s a_R \left( R_y^{-1} a - \frac{\lambda a_{ab}}{\lambda b} R_y b \right), \]

(23)

such that the filter vectors \( w_{L,\text{MWFS-IR}} \) and \( w_{R,\text{MWFS-IR}} \) are parallel, namely, \( w_{L,\text{MWFS-IR}} = (\text{RTF}^x_{\text{in}})^{-1} w_{R,\text{MWFS-IR}} \), and the RTF of the target at the output of the MWF-IR is equal to the input RTF.

For \( \eta R \) greater than zero, the filter vectors of the MWF-IR in (23) can be decomposed as a combination of two beamformers, i.e.,

\[ w_{L,\text{MWFS-IR}} = w_{L,\text{MWFS-IR0}} + \eta R w_{L,\text{MVDR-U}}, \]

\[ w_{R,\text{MWFS-IR}} = w_{R,\text{MWFS-IR0}} + \eta R w_{R,\text{MVDR-U}}, \]

(25)

(26)

where on the one hand, \( w_{L,\text{MWFS-IR0}} \) and \( w_{R,\text{MWFS-IR0}} \) are the filter vectors of the MWF-IR0, and on the other hand, \( w_{L,\text{MVDR-U}} \) and \( w_{R,\text{MVDR-U}} \) are the filter vectors of a minimum variance distortionless response (MVDR) steered towards the interferer, which is denoted as the MVDR-U beamformer, i.e.,

\[ w_{L,\text{MVDR-U}} = \frac{R_y^{-1} b}{\lambda b}, \]

\[ w_{R,\text{MVDR-U}} = \frac{R_{yL}^{-1} b}{\lambda b}, \]

(27)

such that the filter vectors \( w_{L,\text{MVDR-U}} \) and \( w_{R,\text{MVDR-U}} \) are parallel, namely, \( w_{L,\text{MVDR-U}} = (\text{RTF}^x_{\text{in}})^{-1} w_{R,\text{MVDR-U}} \), and the RTF of the interferer at the output of the MVDR-U is equal to the input RTF. The MWF-IR0 is steering a null toward the interferer. However, when RTF estimation errors occur, the interferer is not entirely suppressed and the residual interference will also be perceived as arriving from the target.
direction. To resolve this issue, it is required to set the scaling parameter \( \eta_R \) to a value greater than zero.

In general, the RTF of the target at the output of the MWF-IR is equal to [15]

\[
\text{RTF}_{\text{out}}^T = \frac{w_T^H a}{w_T^H a} = \frac{P_L \lambda_L \left( a_L(1 - \Lambda) + \eta_R b_L \frac{\lambda_L}{P_L \lambda_L} \right)}{P_R \lambda_R \left( a_R(1 - \Lambda) + \eta_R b_R \frac{\lambda_R}{P_R \lambda_R} \right)},
\]

with \( \lambda_L = a^H \mathbf{R}_L^{-1} a \) and \( 0 \leq \Lambda = \frac{|\lambda_{ak}|^2}{\sigma_a^2} \leq 1 \). For \( \eta_R \) greater than zero, since the MWF-IR satisfies the constraints in (22) for the interferer, the output RTF of the interferer is equal to the input RTF, i.e.,

\[
\text{RTF}_{\text{out}}^i = \frac{w_T^H b}{w_T^H b} = \frac{\eta_R b_L}{\eta_R b_R} = \text{RTF}_{\text{in}}^i.
\]

Hence, the MWF-IR preserves the binaural cues of the interferer. For \( \eta_R = 0 \), the output RTF of the interferer is undefined.

It was shown in [13] that the output SINR for the MWF-IR is equal to

\[
\text{SINR}_{\text{MWF-IR}}^{\text{out}} = \varrho_R \text{SINR}_{\text{MWF}}^{\text{out}},
\]

with

\[
\varrho_R = \frac{1 + \Lambda^2 - 2 \Lambda}{1 + \nu \Lambda^2 - 2 \Lambda}, \quad \nu = \left( \frac{\mu + \rho}{\mu \Sigma} \right)^2 - \rho^2 + 2 \mu \rho,
\]

with \( 0 \leq \Sigma = \frac{|\sigma_{ak}|^2}{\sigma_a^2 \sigma_b} \leq 1 \), and \( \sigma_b = b^H \mathbf{R}_b^{-1} b \). Since \( \varrho_R \) is always smaller than or equal to one [13], the output SINR of the MWF-IR is always smaller than or equal to the output SINR of the binaural MWF. Hence,

\[
\text{SINR}_{\text{MWF-IR}}^{\text{out}} \geq \text{SINR}_{\text{MWF-IR}}^{\text{out}_0}, \quad \text{MWF-IR}
\]

IV. EXPERIMENTAL STUDY

In this section, we present simulation results for comparing the performance of the binaural MWF, MWF-N, and the MWF-IR using Behind-The-Ear Impulse Responses (BTE-IR) measured in a reverberant office environment, as described in [17]. To verify the theoretical analysis, we use measured ATFs (a and b) and artificial signals, hence circumventing estimation error issues. All experiments were conducted using \( M = 3 \) microphones, i.e., two microphones on the left hearing aid and one microphone on the right hearing aid, at a sampling frequency of 16 kHz. The acoustic scenario comprised one target at \( \theta_t = 10^\circ \) and 1 m from the artificial head, one interferer at different angles\(^2\) (also 1 m from the artificial head), and diffuse background noise. The angle \( \theta = 0^\circ \) corresponds to the frontal direction and \( \theta = 90^\circ \) corresponds to the right side of the head. The reverberation time is approximately 300 ms. The PSDs of the target and the interferer \( P_s \) and \( P_u \) were calculated from two different speech signals (Welch method using an FFT size of 512 and a Hann window). For the background noise a cylindrically isotropic noise field was assumed, where

\(^2\)Note that the interferer at \( 10^\circ \) was not evaluated.
for all interferer positions. The MWF-IR perfectly preserves the binaural cues of the interferer.

The global output SINR is depicted in Fig. 1(e). It is observed that the global output SINR is lower for both the MWF-N and the MWF-IR as compared to the global output SINR for the binaural MWF.

V. DISCUSSION AND CONCLUSIONS

In this paper, two extensions of the binaural MWF algorithm, namely, the MWF-N and the MWF-IR, were explored. Both the MWF-N and the MWF-IR incorporate binaural cue preservation of the interferer in the binaural MWF cost function. However, their properties are different.

In the MWF-N, the modification of the binaural MWF cost function is for the overall noise signal. The MWF-N filter vectors can be decomposed to the weighted sum of the binaural MWF filter vectors (multiplied by $1 - \eta_N$) and the noisy reference microphone indicator input vectors (multiplied by $\eta_N$). The binaural cues of the target are preserved. However, the parameter $\eta_N$ trades-off the noise reduction and the interferer binaural cue preservation.

In the MWF-IR, a hard constraint is added to the binaural MWF cost function regarding the ATF of the interferer rather than the signal itself. The MWF-IR filter vectors can be decomposed to the sum of the MWF-IR$_0$ filter vectors and MVDR-U filter vectors directed toward the interferer direction (multiplied by $\eta_R$). The MWF-IR$_0$ filter vectors direct a null toward the interferer direction, and the binaural cues of the target are imposed at the output signal, such that the binaural cues of the target are preserved. In order to mask residual interference due to estimation errors and to control the amount of interference reduction, $\eta_R$ should be set to a value greater than zero. Since the MVDR-U preserves the binaural cues of the interferer, the binaural cues of the interferer at the output of the MWF-IR are preserved. The parameter $\eta_R$ trades-off the interference reduction, estimation error, and interferer binaural cue preservation. However, since the binaural cues of the target at the output of the MVDR-U are equal to the binaural cues of the interferer, for $\eta_R$ greater than zero, the binaural cues of the target at the output of the MWF-IR are slightly distorted.

To implement the MWF-N, it is necessary to estimate the spatial correlation matrices of the noisy and the noise-only microphone signals (similar to the implementation requirement for the binaural MWF). To implement the MWF-IR, in addition, it is necessary to estimate the RTF vectors of the interferer (ATF vectors normalized by the ATFs of the reference microphones).

In general, there is a degradation in the performance in terms of SINR for both algorithms as compared to the binaural MWF, as a result of the additional requirement of the binaural cue preservation of the interferer.

ACKNOWLEDGMENT

This work was supported by a joint Lower Saxony-Israeli Project, funded by the State of Lower Saxony in Germany.

REFERENCES