ALTERNATIVE FORMULATION AND ROBUSTNESS ANALYSIS OF THE MULTICHANNEL WIENER FILTER FOR SPATIALLY DISTRIBUTED MICROPHONES

Toby Christian Lawin-Ore[1], Sebastian Stenzel[2], Jürgen Freudenberger[2], Simon Doclo[1]

[1] University of Oldenburg, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Germany
[2] HTWG Konstanz, Institute for System Dynamics, Germany
toby.chris.lawin.ore@uni-oldenburg.de

ABSTRACT

The multichannel Wiener filter (MWF) is a well-known multi-microphone noise reduction technique, which aims to estimate the speech component in one of the microphone signals. Assuming a single speech source, the rank-one property of the speech correlation matrix can be exploited to derive the so-called rank-one MWF (R1-MWF). In this paper, we present an alternative formulation of the MWF (A-MWF), which exploits the assumed rank-one property of the speech correlation matrix in a different way as the R1-MWF. Furthermore, we present a theoretical robustness analysis of the different MWF formulations in presence of spatially white noise. Experimental results show that similarly to the R1-MWF, the proposed A-MWF is less sensitive to estimation errors of the speech correlation matrix and yields a higher output SNR than the standard MWF.

Index Terms— Multichannel Wiener filter, spatially distributed microphones, robustness

1. INTRODUCTION

In speech enhancement applications, the multichannel Wiener filter (MWF), producing a Minimum Mean Square Error (MMSE) estimate of the speech component in one of the microphone signals (referred to as the reference microphone), is widely used to reduce noise and thus improve signal quality [1, 2]. In addition to the standard formulation of the MWF (S-MWF), in the case of a single speech source, a different formulation of the MWF referred to as the rank-one MWF (R1-MWF) can be derived by exploiting the rank-one property of the speech correlation matrix [3, 4]. Basically, for a rank-one speech correlation matrix, the S-MWF and the R1-MWF lead to the same optimal filter coefficients. In practice, both estimates of the S-MWF and the R1-MWF are computed using the speech and noise correlation matrices and it is well known that estimation errors of the speech correlation matrix degrade the performance of both MWF formulations [4]. However, it has been shown in [4] that the R1-MWF is less sensitive to estimation errors and thus leads to a higher output SNR than the S-MWF.

In this paper, we present an alternative formulation of the MWF (A-MWF) for a single speech source, which also exploits the assumed rank-one property of the speech correlation matrix. Furthermore, the theoretical robustness analysis of the S-MWF, the R1-MWF and the A-MWF against small estimation errors shows that, in the presence of spatially white noise, the filter coefficients obtained using the different MWF formulations are, up to a scaling factor, equal to the optimal filter coefficients, i.e., the filter coefficients computed using the true rank-one speech correlation matrix, plus different weighted versions of a bias term. It is shown that the A-MWF puts less weight on the bias term than the R1-MWF, which in turn has a smaller weighted bias term than the S-MWF, such that it can be expected that estimation errors of the speech correlation matrix will have less impact on the performance of the A-MWF than on the R1-MWF and the S-MWF. Experimental results show that the proposed A-MWF leads to a similar output SNR as the R1-MWF and both the A-MWF and the R1-MWF yield a higher output SNR than the S-MWF, especially for spatially distributed microphones.

This work was partly supported by the Research Unit FOR 1732 "Individualized Hearing Acoustics" and the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG).

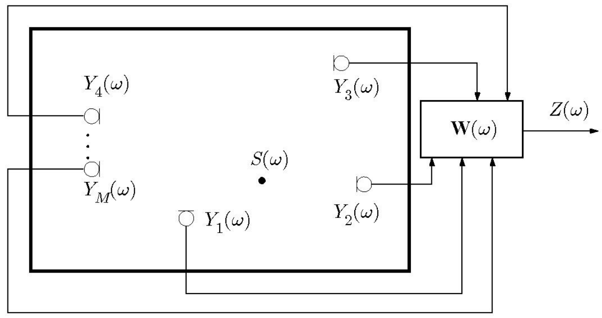

Fig. 1. Configuration of a sensor network with M microphones.

2. SIGNAL MODEL

Consider the acoustical scenario depicted in Figure 1 with M spatially distributed microphones. The received microphone signals can be described in the frequency-domain as

$$\mathbf{Y}(\omega) = \mathbf{A}(\omega)S(\omega) + \mathbf{V}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega), \quad (1)$$

where $\mathbf{Y}(\omega) = [Y_1(\omega) \cdots Y_M(\omega)]^T$ denotes the stacked vector of the microphone signals, $\mathbf{A}(\omega) = [A_1(\omega) \cdots A_M(\omega)]^T$ denotes the stacked vector of the acoustic transfer functions (ATFs) between the speech source $S(\omega)$ and the microphone array, and $\mathbf{X}(\omega)$ and $\mathbf{V}(\omega)$ represent the speech and the noise component in the microphone signals. The output signal $Z(\omega)$ is obtained by filtering and summing the microphone signals, i.e.,

$$Z(\omega) = \mathbf{W}^H(\omega)\mathbf{Y}(\omega) = \mathbf{W}^H(\omega)\mathbf{X}(\omega) + \mathbf{W}^H(\omega)\mathbf{V}(\omega), \quad (2)$$

where $\mathbf{W}(\omega) = [W_1(\omega) \cdots W_M(\omega)]^T$ represents the stacked vector of the filter coefficients. For conciseness the frequency-domain variable ω will be omitted where possible in the remainder of this paper.

The noisy speech correlation matrix $\mathbf{\Phi}_y$, the clean speech correlation matrix $\mathbf{\Phi}_x$ and the noise correlation matrix $\mathbf{\Phi}_v$ are defined as

$$\mathbf{\Phi}_y = \mathcal{E}\{\mathbf{Y}\mathbf{Y}^H\},\ \mathbf{\Phi}_x = \mathcal{E}\{\mathbf{X}\mathbf{X}^H\},\ \mathbf{\Phi}_v = \mathcal{E}\{\mathbf{V}\mathbf{V}^H\}, \tag{3}$$

where $\mathcal{E}\{\cdot\}$ denotes the expected value operator. Assuming that the speech and the noise components are uncorrelated, the correlation matrix $\mathbf{\Phi}_y$ can be expressed as $\mathbf{\Phi}_y = \mathbf{\Phi}_x + \mathbf{\Phi}_v$. For a single speech source, the speech correlation matrix $\mathbf{\Phi}_x$ is a rank-one matrix and is equal to

$$\mathbf{\Phi}_x = \phi_s \mathbf{A}\mathbf{A}^H, \tag{4}$$

with $\phi_s = \mathcal{E}\{|S|^2\}$ the power spectral density (PSD) of the signal S. We assume that a robust voice activity detector is available, such that the correlation matrix $\mathbf{\Phi}_y$ can be estimated during speech-and-noise periods, while the noise correlation matrix $\mathbf{\Phi}_v$ can be estimated during noise-only periods.

3. MULTICHANNEL WIENER FILTERING

The multichannel Wiener filter (MWF) produces a MMSE estimate of the speech component X_{m_0} of the m_0-th microphone, arbitrarily selected to be the reference microphone. To provide a trade-off between speech distortion and noise reduction, the speech-distortion-weighted multichannel Wiener filter has been proposed in [1] and [5], minimizing the weighted sum of the residual noise energy and the speech distortion energy, i.e.,

$$\xi(\mathbf{W}) = \mathcal{E}\{\left|X_{m_0} - \mathbf{W}^H\mathbf{X}\right|^2\} + \mu\mathcal{E}\{|\mathbf{W}^H\mathbf{V}|^2\}, \tag{5}$$

where μ is a trade-off parameter between noise reduction and speech distortion. The S-MWF minimizing the cost function in (5) is given by

$$\boxed{\mathbf{W}_{\text{S-MWF}} = (\mathbf{\Phi}_x + \mu\mathbf{\Phi}_v)^{-1}\mathbf{\Phi}_x\mathbf{e}_{m_0}} \tag{6}$$

with $\mathbf{e}_{m_0}$ an M-dimensional vector with the m_0-th element equal to 1 and all other elements equal to 0, i.e., the vector selecting the column that corresponds to the reference microphone.

Using the rank-one property of the speech correlation matrix for a single speech source and applying the matrix inversion lemma, the optimal MWF can be expressed as [2]

$$\mathbf{W}_{\text{opt}} = \frac{\mathbf{\Phi}_v^{-1}\phi_s\mathbf{A}A_{m_0}^*}{\mu + \phi_s\mathbf{A}^H\mathbf{\Phi}_v^{-1}\mathbf{A}}, \tag{7}$$

with A_{m_0} the ATF relating the speech source to the reference microphone. As can be observed, the computation of the optimal MWF in (7) requires knowledge about the ATFs and the PSD ϕ_s of the signal S, which are not available.

On the other hand, using (7) and the fact that $\phi_s\mathbf{A}^H\mathbf{\Phi}_v^{-1}\mathbf{A} = \text{tr}(\mathbf{\Phi}_v^{-1}\mathbf{\Phi}_x)$, the R1-MWF can be derived as [3, 4]

$$\boxed{\mathbf{W}_{\text{R1-MWF}} = \frac{\mathbf{\Phi}_v^{-1}\mathbf{\Phi}_x\mathbf{e}_{m_0}}{\mu + \text{tr}(\mathbf{\Phi}_v^{-1}\mathbf{\Phi}_x)}} \tag{8}$$

where $\text{tr}(\cdot)$ denotes the trace operator.

As an alternative to the approach used to derive the R1-MWF, we propose to use the rank-one property of the speech correlation matrix in a different way by reformulating the correlation vector $\mathbf{\Phi}_x\mathbf{e}_{m_0}$ in (6) as

$$\mathbf{\Phi}_x\mathbf{e}_{m_0} = \phi_s\mathbf{A}A_{m_0}^* = \phi_s\mathbf{A}\mathbf{A}^H\frac{\phi_s\mathbf{A}A_{m_0}^*}{\phi_s\mathbf{A}^H\mathbf{A}} = \mathbf{\Phi}_x\mathbf{g}, \tag{9}$$

with

$$\mathbf{g} = \frac{\phi_s\mathbf{A}A_{m_0}^*}{\phi_s\mathbf{A}^H\mathbf{A}} = \frac{\mathbf{\Phi}_x\mathbf{e}_{m_0}}{\text{tr}(\mathbf{\Phi}_x)}. \tag{10}$$

Using (10), the A-MWF is then given by

$$\boxed{\mathbf{W}_{\text{A-MWF}} = (\mathbf{\Phi}_x + \mu\mathbf{\Phi}_v)^{-1}\mathbf{\Phi}_x\mathbf{g}} \tag{11}$$

It should be noted that the MWF formulation in (8) and (11) have been derived assuming a rank-one speech correlation matrix. For the case that the speech correlation matrix is a rank-one matrix, (6), (8) and (11) are the same, i.e., equal to the optimal MWF in (7). However, the MWF formulations in (6), (8) and (11) can also be used when $\mathbf{\Phi}_x$ is not a rank-one matrix (e.g. occurring when estimating $\mathbf{\Phi}_x$ in practice) leading to different filter coefficients with different performance. In the next section, a theoretical analysis of the influence of estimation errors on the performance of the different MWF formulations will be presented.

4. THEORETICAL ROBUSTNESS ANALYSIS

In practice, the speech correlation matrix $\mathbf{\Phi}_x$ is typically estimated as

$$\hat{\mathbf{\Phi}}_x = \hat{\mathbf{\Phi}}_y - \hat{\mathbf{\Phi}}_v, \tag{12}$$

where $\hat{\mathbf{\Phi}}_y$ and $\hat{\mathbf{\Phi}}_v$ are estimates of the noisy speech and noise correlation matrices, respectively. Unfortunately, using (12) there is no guaranty that the estimated speech correlation matrix $\hat{\mathbf{\Phi}}_x$ has the previously assumed rank-one property. Hence, by plugging $\hat{\mathbf{\Phi}}_x$ and $\hat{\mathbf{\Phi}}_v$ into (6), (8) and (11), different filters are obtained resulting in different performance. We will now analyze the impact of estimation errors on the different MWF formulations.

For the robustness analysis, we assume a spatially white noise field, i.e., $\mathbf{\Phi}_v = \phi_v\mathbf{I}$ with ϕ_v the noise PSD. Although this is a very restrictive assumption, it is used to theoretically demonstrate the robustness of the different formulations of the MWF against estimation errors. Furthermore, we suppose that in case of a spatially white noise field, the estimated speech correlation matrix $\hat{\mathbf{\Phi}}_x$ can be modeled using the theoretical rank-one correlation matrix $\mathbf{\Phi}_x = \phi_s\mathbf{A}\mathbf{A}^H$ as

$$\hat{\mathbf{\Phi}}_x = \mathbf{\Phi}_x + \beta\mathbf{I}, \tag{13}$$

where $\beta\mathbf{I}$ represents the estimation errors and β is assumed to be a positive scaling factor. Moreover, we assume that only small estimation errors occur, such that the matrix $\mu\phi_v\mathbf{I}$ is *dominant* compared to the matrix $\beta\mathbf{I}$, i.e., $\beta \ll \mu\phi_v$.

4.1. Analysis of the standard MWF

By plugging (13) into (6), the estimated S-MWF can be expressed as

$$\hat{\mathbf{W}}_{\text{S-MWF}} = (\mathbf{\Phi}_x + \beta\mathbf{I} + \mu\phi_v\mathbf{I})^{-1}(\mathbf{\Phi}_x + \beta\mathbf{I})\mathbf{e}_{m_0}. \tag{14}$$

Using the fact that $\beta \ll \mu\phi_v$ and applying the matrix inversion lemma, (14) can be rewritten as

$$\begin{aligned}\hat{\mathbf{W}}_{\text{S-MWF}} &= (\mathbf{\Phi}_x + \mu\phi_v\mathbf{I})^{-1}\big(\mathbf{\Phi}_x + \beta\mathbf{I}\big)\mathbf{e}_{m_0} \qquad (15)\\ &= \left[\frac{1}{\mu\phi_v}\mathbf{I} - \frac{\frac{1}{\mu\phi_v}\mathbf{I}\phi_s\mathbf{A}\mathbf{A}^H\frac{1}{\mu\phi_v}\mathbf{I}}{1 + \frac{1}{\mu\phi_v}\phi_s\mathbf{A}^H\mathbf{A}}\right]\big(\mathbf{\Phi}_x + \beta\mathbf{I}\big)\mathbf{e}_{m_0}\\ &= \frac{\mu\phi_v - \beta}{\mu\phi_v}\left[\frac{\phi_s\mathbf{A}A_{m_0}^*}{\mu\phi_v + \phi_s\mathbf{A}^H\mathbf{A}}\right] + \frac{\beta}{\mu\phi_v}\mathbf{e}_{m_0}\\ &\approx \mathbf{W}_{\text{opt}} + \mathbf{W}_{\text{S-MWF}}^{\text{err}},\end{aligned}$$

with

$$\boxed{\mathbf{W}_{\text{S-MWF}}^{\text{err}} = \frac{\beta}{\mu\phi_v}\mathbf{e}_{m_0}} \tag{16}$$

Hence, in case of estimation errors, the computed S-MWF is equal to the optimal MWF plus an error vector $\mathbf{W}^{\text{err}}_{\text{S-MWF}}$. As can be observed, the error vector is a scaled version of the trivial filter $\mathbf{e}_{m_0}$, i.e., that a scaled version of the noisy reference microphone signal is added to the output signal of the optimal MWF. Therefore, it can be expected that this will lead to a performance degradation, e.g., to a poor output SNR compared to the optimal filter.

4.2. Analysis of the rank-one MWF

By plugging (13) into (8), the R1-MWF can be rewritten as

$$\begin{aligned}\hat{\mathbf{W}}_{\text{R1-MWF}} &= \frac{(\phi_v\mathbf{I})^{-1}(\mathbf{\Phi}_x+\beta\mathbf{I})\mathbf{e}_{m_0}}{\mu+\text{tr}((\phi_v\mathbf{I})^{-1}(\mathbf{\Phi}_x+\beta\mathbf{I}))} \quad (17)\\ &= \frac{\phi_s\mathbf{A}A^*_{m_0}}{\mu\phi_v+\text{tr}(\mathbf{\Phi}_x)+M\beta}+\frac{\beta\mathbf{e}_{m_0}}{\mu\phi_v+\text{tr}(\mathbf{\Phi}_x)+M\beta}\\ &= \frac{\mu\phi_v+\text{tr}(\mathbf{\Phi}_x)}{\mu\phi_v+\text{tr}(\mathbf{\Phi}_x)+M\beta}\Big[\mathbf{W}_{\text{opt}}+\mathbf{W}^{\text{err}}_{\text{R1-MWF}}\Big],\end{aligned}$$

with

$$\boxed{\mathbf{W}^{\text{err}}_{\text{R1-MWF}}=\frac{\beta}{\mu\phi_v+\text{tr}(\mathbf{\Phi}_x)}\mathbf{e}_{m_0}} \quad (18)$$

Similarly to the S-MWF, the estimated R1-MWF is, up to a scaling factor, equal to the optimal MWF plus a scaled version of the trivial filter $\mathbf{e}_{m_0}$. By comparing (16) and (18), it can be easily observed that the error vector $\mathbf{W}^{\text{err}}_{\text{S-MWF}}$ is parallel to the error vector $\mathbf{W}^{\text{err}}_{\text{R1-MWF}}$. Furthermore, the following inequality holds, i.e.,

$$\frac{\beta}{\mu\phi_v+\text{tr}(\mathbf{\Phi}_x)}\leq\frac{\beta}{\mu\phi_v}. \quad (19)$$

Hence, $\mathbf{W}^{\text{err}}_{\text{R1-MWF}}$ has a smaller amplitude than $\mathbf{W}^{\text{err}}_{\text{S-MWF}}$, which means that the estimation errors in the speech correlation matrix have less effect on the estimated R1-MWF than on the S-MWF and it can be expected that the R1-MWF will lead to less performance degradation than the S-MWF.

4.3. Analysis of the alternative formulation of MWF

By plugging (13) into (10) and (11), the A-MWF can be rewritten as

$$\begin{aligned}\hat{\mathbf{W}}_{\text{A-MWF}} &= (\mathbf{\Phi}_x+\beta\mathbf{I}+\mu\phi_v\mathbf{I})^{-1}\frac{\big(\mathbf{\Phi}_x+\beta\mathbf{I}\big)^2}{\text{tr}\big(\mathbf{\Phi}_x+\beta\mathbf{I}\big)}\mathbf{e}_{m_0} \quad (20)\\ &= (\mathbf{\Phi}_x+\mu\phi_v\mathbf{I})^{-1}\frac{\text{tr}(\mathbf{\Phi}_x)\mathbf{\Phi}_x+\beta^2\mathbf{I}+2\beta\mathbf{\Phi}_x}{\text{tr}(\mathbf{\Phi}_x)+M\beta}\mathbf{e}_{m_0}\\ &= \frac{\text{tr}(\mathbf{\Phi}_x)+2\beta}{\text{tr}(\mathbf{\Phi}_x)+M\beta}(\mathbf{\Phi}_x+\mu\phi_v\mathbf{I})^{-1}(\mathbf{\Phi}_x+\gamma\mathbf{I})\mathbf{e}_{m_0},\end{aligned}$$

with

$$\gamma=\frac{\beta^2}{\text{tr}(\mathbf{\Phi}_x)+2\beta}\leq\frac{\beta}{2}\ll\mu\phi_v. \quad (21)$$

Similarly as in (15), (20) can be rewritten as

$$\begin{aligned}\hat{\mathbf{W}}_{\text{A-MWF}} &= \frac{\text{tr}(\mathbf{\Phi}_x)+2\beta}{\text{tr}(\mathbf{\Phi}_x)+M\beta}\left[\frac{\phi_s\mathbf{A}A^*_{m_0}}{\mu\phi_v+\phi_s\mathbf{A}^H\mathbf{A}}+\frac{\gamma}{\mu\phi_v}\mathbf{e}_{m_0}\right] \quad (22)\\ &\approx \frac{\text{tr}(\mathbf{\Phi}_x)+2\beta}{\text{tr}(\mathbf{\Phi}_x)+M\beta}\Big[\mathbf{W}_{\text{opt}}+\mathbf{W}^{\text{err}}_{\text{A-MWF}}\Big],\end{aligned}$$

with

$$\boxed{\mathbf{W}^{\text{err}}_{\text{A-MWF}}=\frac{\gamma}{\mu\phi_v}\mathbf{e}_{m_0}} \quad (23)$$

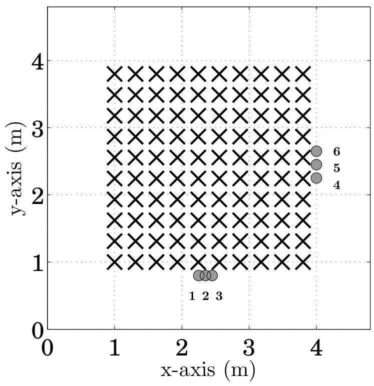

Fig. 2. The scenario of an acoustic sensor network with $M = 6$ microphones.

As can be observed, the resulting filter corresponds, up to a scaling factor, to the optimal MWF plus an error vector $\mathbf{W}^{\text{err}}_{\text{A-MWF}}$, which is parallel to the error vectors $\mathbf{W}^{\text{err}}_{\text{S-MWF}}$ and $\mathbf{W}^{\text{err}}_{\text{R1-MWF}}$. For small estimation errors in the speech correlation matrix, i.e., $\beta\ll\mu\phi_v$, it can be shown that the following inequality holds, i.e.,

$$\frac{\gamma}{\mu\phi_v}=\frac{\beta^2}{\mu\phi_v(\text{tr}(\mathbf{\Phi}_x)+2\beta)}\leq\frac{\beta}{\mu\phi_v+\text{tr}(\mathbf{\Phi}_x)}\leq\frac{\beta}{\mu\phi_v}. \quad (24)$$

Hence, the estimation errors in the speech correlation matrix will have less effect on the estimated A-MWF than on both estimates of the R1-MWF and S-MWF, and it can be expected that the A-MWF will lead to less performance degradation than the R1-MWF and the S-MWF.

5. EXPERIMENTAL RESULTS

In this section we investigate the influence of estimation errors on the performance of the MWF using the different MWF formulations discussed in Section 3 .

5.1. Setup and performance measures

In a room with dimensions 4.8m×4.8m×3m and $T_{60} = 400$ms, we consider the acoustic scenario depicted in Figure 2 with 6 spatially distributed microphones and a single desired speech source. The circles represent the microphone positions and the cross markers various positions of the desired source. The desired signal has been generated by convolving a clean speech signal from the HINT-database [6] with impulse responses simulated using the image model [7, 8]. The sampling frequency is $f_s = 16$kHz.

For each position of the desired source, the theoretical rank-one speech correlation matrix $\mathbf{\Phi}_x$ is calculated by plugging into (4), the ATF $\mathbf{A}$ computed using the simulated room impulse responses and the source PSD ϕ_s estimated using the clean speech signal. The a-priori input SNR $\frac{\phi_s}{\phi_v}$ is set to 5 dB.

In our STFT-based implementation, we have used the overlap/add method with a Hann analysis and synthesis window, and 50% overlap. The used FFT length is $N_{\text{FFT}} = 1024$. The different MWF formulations presented in Section 3 are computed for each frequency bin. Using a perfect voice activity detector, the correlation matrices $\mathbf{\Phi}_y(\omega)$ and $\mathbf{\Phi}_v(\omega)$ are estimated in batch mode by using all speech + noise frames and all noise-only frames respectively, i.e.,

$$\hat{\mathbf{\Phi}}_y(\omega)=\frac{1}{F_x}\sum_{F_x}\mathbf{Y}(\omega)\mathbf{Y}^H(\omega), \quad (25)$$

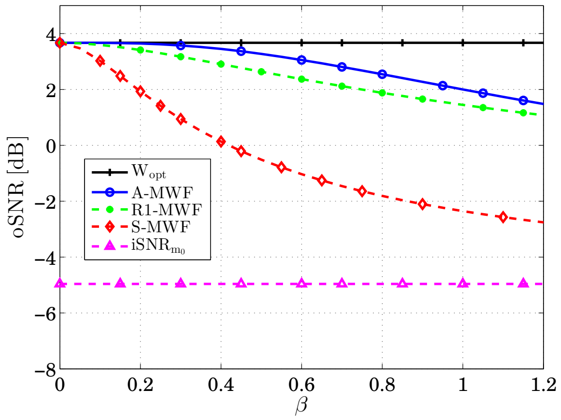

Fig. 3. Influence of estimation errors on the narrowband output SNR of the MWF filters ($f = 1500$ Hz).

$$\hat{\mathbf{\Phi}}_v(\omega) = \frac{1}{F_v} \sum_{F_v} \mathbf{V}(\omega)\mathbf{V}^H(\omega), \tag{26}$$

where F_x and F_v are the number of frames during speech + noise and noise-only periods. The trade-off parameter is set to $\mu = 1$ and the first microphone is selected as the reference microphone, i.e., $m_0 = 1$.

To evaluate the performance of the MWF filters, we consider the narrowband output SNR which is defined as

$$\text{oSNR}(\omega) = \frac{\hat{\mathbf{W}}^H(\omega)\mathbf{\Phi}_x(\omega)\hat{\mathbf{W}}(\omega)}{\hat{\mathbf{W}}^H(\omega)\mathbf{\Phi}_v(\omega)\hat{\mathbf{W}}(\omega)}. \tag{27}$$

Furthermore, we define the broadband output SNR of the MWF as

$$\text{oSNR}^B = \frac{\sum_\omega \hat{\mathbf{W}}^H(\omega)\mathbf{\Phi}_x(\omega)\hat{\mathbf{W}}(\omega)}{\sum_\omega \hat{\mathbf{W}}^H(\omega)\mathbf{\Phi}_v(\omega)\hat{\mathbf{W}}(\omega)}. \tag{28}$$

5.2. Robustness analysis

To analyze the theoretical robustness of the MWF filters in case of estimation errors, we first simulate the estimated speech correlation matrix as the sum of the true rank-one speech correlation matrix and a normalized error matrix, i.e.,

$$\hat{\mathbf{\Phi}}_x(\omega) = \phi_s(\omega)\mathbf{A}(\omega)\mathbf{A}^H(\omega) + \beta||\mathbf{\Phi}_v(\omega)||\mathbf{I}, \tag{29}$$

where $\mathbf{A}$ has been computed for a desired source located at the position with coordinates (1.2, 2.8). For this experiment a simulated spatially white noise field has been used. The estimates of the different MWF Formulations have been computed by plugging (26) and (29) into (6), (8) and (11). At frequency $f = 1500$ Hz, Figure 3 depicts the narrowband output SNR of the S-MWF, the R1-MWF and the A-MWF as a function of β . As can be seen, when β increases, the narrowband output SNRs of the different MWF formulations decrease. However, for small estimation errors, the narrowband output SNRs obtained using the A-MWF and the R1-MWF are close to the optimal output SNR, i.e., the output SNR of the MWF in absence of estimation errors. Furthermore, the A-MWF and the R1-MWF yield a higher narrowband output SNR than the S-MWF, especially for large β. Moreover, as can be observed, the output SNR of the S-MWF converges to the input SNR of the reference microphone for very large estimation errors. Similar results are obtained for other frequencies.

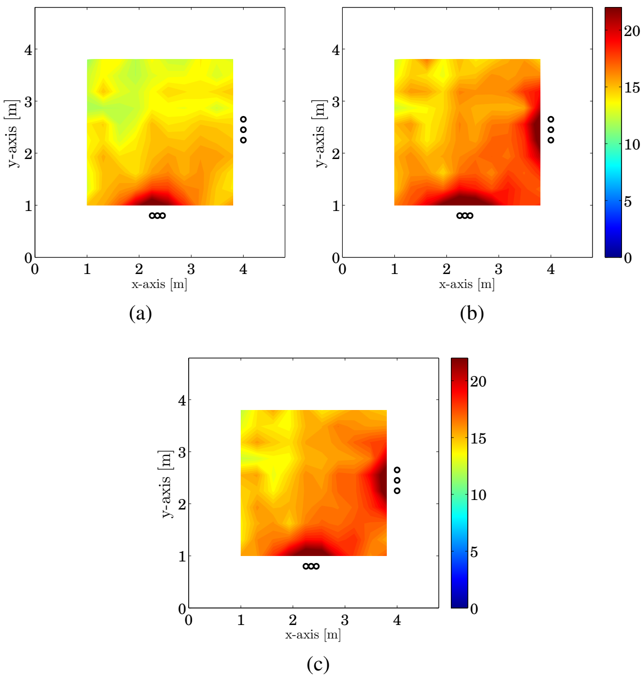

Fig. 4. Position-dependent broadband output SNR of the different MWF filters: (a) S-MWF; (b) R1-MWF; (c) A-MWF.

5.3. Performance for various source positions

In the second experiment, a simulated diffuse noise field has been used and the speech correlation matrix has been estimated using (12), (25) and (26). Figures 4(a), 4(b) and 4(c) show the broadband output SNR of the MWF for different positions of the desired source for the S-MWF, the R1-MWF and the A-MWF, respectively. First, it can be observed that the S-MWF leads to good results at some positions of the desired source but due to estimation errors in the speech correlation matrix to poor results at other positions. For example, a relatively small output SNR is achieved when the speaker is located in the area close to the microphones 4 to 6. In contrast to the S-MWF, the R1-MWF and the A-MWF lead to higher broadband output SNRs than the S-MWF at all positions of the desired source. Furthermore, by comparing Figure 4(b) and Figure 4(c), it can be seen that the A-MWF and the R1-MWF yield similar results at all positions of the desired source.

6. CONCLUSION

In this paper, an alternative formulation of the MWF (A-MWF), which exploits the rank-one property of the speech correlation matrix for a single speech source, has been presented. The theoretical robustness analysis of the standard MWF (S-MWF), the rank-one MWF (R1-MWF) and the A-MWF in presence of spatially white noise has shown that in case of estimation errors in the speech correlation matrix, the A-MWF leads to a higher narrowband output SNR than the R1-MWF and the S-MWF. Moreover, simulation results using a diffuse noise field have shown that the A-MWF and the R1-MWF yield similar broadband output SNR but lead to higher broadband output SNR than the S-MWF, especially in the context of acoustic sensor networks.

7. REFERENCES

[1] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication, special issue on Speech Enhancement*, vol. 49, pp. 636–656, July 2007.

[2] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, ch. 9, pp. 269–302, Wiley, 2010.

[3] M. Souden, J. Benesty, and S. Affes, "On Optimal Frequency-Domain Multichannel Linear Filtering for Noise Reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, pp. 260–276, Feb. 2010.

[4] B. Cornelis and M. Moonen, "Performance analysis of multichannel Wiener filter based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 1368–1381, July 2011.

[5] S. Doclo, T. van den Bogaert, J. Wouters, and M. Moonen, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, pp. 38–51, Jan. 2009.

[6] M. Nilsson, S. D. Soli, and A. Sullivan, "Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise," *Journal of the Acoustical Society of America*, vol. 95, pp. 1085–1099, Feb. 1994.

[7] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, June 1979.

[8] E. A. P. Habets, "Room impulse response (RIR) generator," *available: http://home.tiscali.nl/ehabets/rir generator.html*, 2008.