USING STATISTICAL ROOM ACOUSTICS FOR ANALYSING THE OUTPUT SNR OF THE MWF IN ACOUSTIC SENSOR NETWORKS

Toby Christian Lawin-Ore, Simon Doclo

University of Oldenburg, Institute of Physics, Signal Processing Group, Oldenburg, Germany
{toby.chris.lawin.ore, simon.doclo}@uni-oldenburg.de

ABSTRACT

In the context of *acoustic sensor networks* with spatially distributed microphones, the selection of the subset of microphones yielding the best performance is of great interest. Subset selection can be achieved by comparing the theoretical performance of different subsets of microphones. In this paper, we derive an analytical expression for the spatially averaged output SNR of the multi-channel Wiener filter (MWF) in a diffuse noise field, exploiting the statistical properties of the acoustic transfer functions (ATFs) between the desired source and the microphones. This analytical expression only requires the room properties and the source-microphone distances to be known. Simulation results show that the spatially averaged output SNR obtained using the statistical properties of ATFs is similar to the average output SNR obtained using simulated ATFs, therefore providing an efficient way to compare the performance of different subsets of microphones.

Index Terms— Multi-channel Wiener filter, statistical room acoustics, acoustic sensor network

1. INTRODUCTION

For every speech enhancement algorithms it is of significant interest to be able to compute its theoretical performance, e.g. output SNR, for different acoustical scenarios (microphone configuration, source position, noise field). This enables to compare the performance of different microphone configurations such that the microphone configuration yielding the best performance can be selected.

In speech enhancement applications, the multi-channel Wiener filter (MWF) is often used to reduce noise and thus improve signal quality [1]. The MWF performs noise reduction by estimating the desired signal component in one of the microphones, referred to as the reference microphone. In [2] the theoretical performance of the MWF has been analyzed for different noise fields (diffuse and coherent noise sources). It has been shown that the output SNR of the MWF can be computed using the noise correlation matrix and the ATFs between the desired source and the microphones. Hence, for every source-microphones configuration, the theoretical performance of the MWF can be computed using measured or simulated noise correlation matrices and ATFs. If we want to compare the performance for a large number of source-microphones configurations (and assuming that an estimated or simulated noise correlation matrix is available), then either a large number of ATFs need to be measured, which could be very time-consuming, or the performance of the MWF can be *numerically* simulated, by simulating the ATFs using the image model [3] or room acoustics software. Obviously, in the last case the room properties (room dimensions, wall absorption coefficients) and the positions of the source and microphones need to be known.

Statistical room acoustics (SRA) has been used, e.g., to express statistical properties of ATFs [4], and to derive analytical expressions for performance measures. E.g. in [5] the robustness of an equalization technique has been analyzed using SRA. Furthermore, in [6] a method to predict the SNR improvement of a delay-and-sum beamformer with two microphones using the statistical properties of ATFs has been presented. In this paper an (approximate) analytical expression for the output SNR of the MWF in a diffuse noise field is derived using SRA, incorporating statistical properties of the ATFs. The proposed method allows for easy performance comparison of different source positions and subsets of microphones, without having to measure or numerically simulate ATFs.

2. SIGNAL MODEL AND CONFIGURATION

Figure 1 shows the configuration of M microphones located at positions $\mathbf{p}_m = [x_m \;\; y_m \;\; z_m]^T, m = 0 \cdots M-1$, and a single speech source $S(\omega)$ located at position $\mathbf{p}_s = [x_s \;\; y_s \;\; z_s]^T$. The complete microphone array configuration can be described by the $3 \times M$-matrix $\mathbf{P}_{mic} = [\mathbf{p}_0 \cdots \mathbf{p}_{M-1}]$. We define the relative distance between the speech source and the microphones as

$$\mathbf{d} = \begin{bmatrix} d_0 \\ \vdots \\ d_{M-1} \end{bmatrix} = \begin{bmatrix} \|\mathbf{p}_s - \mathbf{p}_0\| \\ \vdots \\ \|\mathbf{p}_s - \mathbf{p}_{M-1}\| \end{bmatrix}. \tag{1}$$

This work was supported by the Research Unit FOR 1732 "Individualized Hearing Acoustics", funded by the German Research Foundation (DFG).

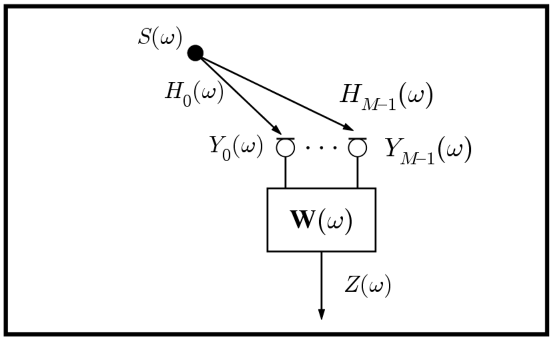

Fig. 1. Configuration of an array with M microphones.

The mth microphone signal $Y_m(\omega)$ can be described in the frequency domain as

$$\begin{aligned} Y_m(\omega) &= H_m(\omega)S(\omega) + V_m(\omega), m = 0 \ldots M-1 \\ &= X_m(\omega) + V_m(\omega), \end{aligned} \tag{2}$$

where $H_m(\omega)$ represents the ATF between the speech source $S(\omega)$ and the mth microphone, and $X_m(\omega)$ and $V_m(\omega)$ represent the speech and the noise component in the mth microphone signal. We define the M-dimensional stacked signal vector $\mathbf{Y}(\omega)$ as

$$\mathbf{Y}(\omega) = \begin{bmatrix} Y_0(\omega) \\ \vdots \\ Y_{M-1}(\omega) \end{bmatrix}, \tag{3}$$

which can be written as

$$\mathbf{Y}(\omega) = \mathbf{X}(\omega) + \mathbf{V}(\omega), \tag{4}$$

where the vectors $\mathbf{X}(\omega)$ and $\mathbf{V}(\omega)$ are similarly defined as $\mathbf{Y}(\omega)$. The output signal $Z(\omega)$ is obtained by filtering and summing the microphone signals, i.e.,

$$\begin{aligned} Z(\omega) &= \mathbf{W}^H(\omega)\mathbf{X}(\omega) + \mathbf{W}^H(\omega)\mathbf{V}(\omega) \\ &= Z_x(\omega) + Z_v(\omega), \end{aligned} \tag{5}$$

where $\mathbf{W}(\omega) = [W_0(\omega) \cdots W_{M-1}(\omega)]^T$ represents the stacked vector of the filter coefficients, and $Z_x(\omega)$ and $Z_v(\omega)$ correspond to the estimated speech and residual noise component respectively.

3. MULTI-CHANNEL WIENER FILTERING

The concept of multi-channel Wiener filtering (MWF) is based on estimating the speech component X_{m_0} of the m_0th microphone, arbitrarily selected as the reference microphone. The MWF produces a minimum-mean-square error (MMSE) estimate by minimizing the MSE cost function [1]

$$\xi(\mathbf{W}(\omega)) = \mathcal{E}\{\left|X_{m_0}(\omega) - \mathbf{W}^H(\omega)\mathbf{Y}(\omega)\right|^2\}, \tag{6}$$

where $\mathcal{E}\{.\}$ denotes the expected value operator. The solution of this minimization problem is given by

$$\mathbf{W}_{m_0}(\omega) = \mathbf{\Phi}_y^{-1}(\omega)\mathbf{\Phi}_x(\omega)\mathbf{e}_{m_0}, \tag{7}$$

with $\mathbf{\Phi}_y(\omega) = \mathcal{E}\{\mathbf{Y}(\omega)\mathbf{Y}^H(\omega)\}$, $\mathbf{\Phi}_x(\omega) = \mathcal{E}\{\mathbf{X}(\omega)\mathbf{X}^H(\omega)\}$ the noisy and clean speech correlation matrix, and $\mathbf{e}_{m_0}$ an M-dimensional vector with the m_0th element equal to 1 and all other elements equal to 0, selecting the column that corresponds to the reference microphone.

Assuming that the speech and the noise components are uncorrelated, the correlation matrix $\mathbf{\Phi}_y(\omega)$ can be expressed as

$$\mathbf{\Phi}_y(\omega) = \mathbf{\Phi}_x(\omega) + \mathbf{\Phi}_v(\omega), \tag{8}$$

where $\mathbf{\Phi}_v(\omega)$ represents the noise correlation matrix, i.e., $\mathbf{\Phi}_v(\omega) = \mathcal{E}\{\mathbf{V}(\omega)\mathbf{V}^H(\omega)\}$, which is assumed to be full rank. Using a robust VAD, the correlation matrix $\mathbf{\Phi}_y(\omega)$ can be estimated during speech + noise periods, while the correlation matrix $\mathbf{\Phi}_v(\omega)$ can be estimated during speech pauses.

For conciseness the frequency-domain variable ω will be omitted where possible in the remainder of this paper.

4. THEORETICAL PERFORMANCE OF MWF

Similarly as in [1], the theoretical performance of the MWF will be presented in this section.

For a single desired speech source, the speech correlation matrix $\mathbf{\Phi}_x$ is a rank-one matrix and is equal to

$$\mathbf{\Phi}_x = \phi_s \mathbf{H}\mathbf{H}^H, \tag{9}$$

with ϕ_s the power spectral density (PSD) of the source S, i.e. $\phi_s = \mathcal{E}\{|S|^2\}$ and $\mathbf{H} = [H_0 \cdots H_{M-1}]^T$ the stacked vector of the ATFs. Using (8) and (9), the correlation matrix $\mathbf{\Phi}_y$ can be written as

$$\mathbf{\Phi}_y = \phi_s \mathbf{H}\mathbf{H}^H + \mathbf{\Phi}_v. \tag{10}$$

Using the matrix inversion lemma, the inverse matrix $\mathbf{\Phi}_y^{-1}$ can be expressed as

$$(\phi_s \mathbf{H}\mathbf{H}^H + \mathbf{\Phi}_v)^{-1} = \Big(\mathbf{I} - \frac{\mathbf{\Phi}_v^{-1}\mathbf{H}\mathbf{H}^H}{\phi_s^{-1} + \Lambda}\Big)\mathbf{\Phi}_v^{-1}, \tag{11}$$

where

$$\boxed{\Lambda = \mathbf{H}^H \mathbf{\Phi}_v^{-1}\mathbf{H}.} \tag{12}$$

Inserting (9) and (11) into (7) yields

$$\boxed{\mathbf{W}_{m_0} = \frac{\mathbf{\Phi}_v^{-1}\mathbf{H}}{\phi_s^{-1} + \Lambda} H_{m_0}^*.} \tag{13}$$

The output SNR of the MWF is equal to

$$\mathrm{SNR}_{\mathrm{out}} = \frac{\mathcal{E}\{|Z_x|^2\}}{\mathcal{E}\{|Z_v|^2\}} = \frac{\mathbf{W}_{m_0}^H \mathbf{\Phi}_x \mathbf{W}_{m_0}}{\mathbf{W}_{m_0}^H \mathbf{\Phi}_v \mathbf{W}_{m_0}}, \tag{14}$$

which using (9) and (13) can be written as

$$\boxed{\mathrm{SNR}_{\mathrm{out}} = \phi_s \Lambda.} \tag{15}$$

If we assume that the noise field is homogeneous [1], i.e. $\mathbf{\Phi}_v(m,m) = \phi_v, \forall m$, then the noise correlation matrix can be expressed as

$$\mathbf{\Phi}_v = \phi_v \mathbf{\Gamma}_v, \tag{16}$$

where ϕ_v denotes the noise PSD and $\mathbf{\Gamma}_v$ denotes the noise coherence matrix. The output SNR of the MWF given by (15) can then be written as

$$\boxed{\mathrm{SNR}_{\mathrm{out}} = \frac{\phi_s}{\phi_v}\rho,} \tag{17}$$

with

$$\boxed{\rho = \mathbf{H}^H \mathbf{\Gamma}_v^{-1} \mathbf{H}.} \tag{18}$$

Hence, the output SNR of the MWF depends on the a priori input SNR $\frac{\phi_s}{\phi_v}$, and on the spatial characteristics ρ, i.e. the ATFs $\mathbf{H}$ between the source and the microphones, and the spatial characteristics of the noise field described by the noise coherence matrix $\mathbf{\Gamma}_v$. Therefore, if $\mathbf{\Gamma}_v$ and $\mathbf{H}$ are known, the output SNR of the MWF can be calculated.

5. PERFORMANCE OF MWF USING STATISTICAL PROPERTIES OF ATFS

The theory of statistical room acoustics is based on the assumption that the phase and the amplitude of reverberant plane waves arriving at a point in a room are close to random. The resulting reverberant sound field can then be considered as uniformly distributed in the entire room [4]. This model of the reverberant sound field is valid only if

1. The dimensions of the room are large relative to the wavelength of the considered signals. For example, in a room with dimensions 8 m×6 m×5 m this condition is satisfied for typical speech applications, where we consider a minimum frequency of 300 Hz (corresponding to a wavelength of about 1.14 m).
2. The normal modes of the room overlap each other at least 3:1. This condition is satisfied for frequencies that exceed the Schroeder frequency $f_g = 2000\sqrt{T_{60}/V}$, where V is the volume of the room under consideration. For example, in a room with dimensions 8 m×6 m×5 m and reverberation time $T_{60} = 0.25$ s, $f_g = 65$ Hz.
3. The microphones and the source are located in the interior of the room at least half wavelength away from the walls. For speech signals with a lower frequency of 300 Hz, the microphones and the source must be at least 0.57 m away from the walls.

[1] The assumption of a homogeneous noise field always holds for a diffuse noise field or when the microphones are closely spaced.

5.1. Statistical properties of ATFs

Without loss of generality, the vector containing the ATFs between the source located at position $\mathbf{p}_s$ and the M microphones located at the positions $\mathbf{p}_m$ can be decomposed as

$$\mathbf{H}(\boldsymbol{\theta}) \;=\; \mathbf{H}_d(\boldsymbol{\theta}) + \mathbf{H}_r(\boldsymbol{\theta}), \tag{19}$$

where $\boldsymbol{\theta} = [\mathbf{p}_s,\ \mathbf{P}_{mic}]$ and $\mathbf{H}_d(\boldsymbol{\theta})$ and $\mathbf{H}_r(\boldsymbol{\theta})$ are the vectors corresponding to the direct and the reverberant component of the ATFs.

We define the spatial expectation operator $\mathcal{E}_{\boldsymbol{\theta}}\{\cdot\}$ as the ensemble average over all realizations of $\boldsymbol{\theta}$. Using SRA, the following statistical properties of ATFs are then given [4]:

- A1 For a fixed relative distance $\mathbf{d}$ between source and microphones, the direct path components are independent of the realization of $\boldsymbol{\theta}$, i.e.,

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})H^*_{n,d}(\boldsymbol{\theta})|d_m,d_n\} = \frac{e^{j\frac{\omega}{c}(d_n-d_m)}}{(4\pi)^2 d_m d_n}\ \forall m,n. \tag{20}$$

- A2 The spatially expected correlation between the reverberant components of the ATF of the mth and the nth microphone is independent of $\mathbf{d}$ and is given by

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,r}(\boldsymbol{\theta})H^*_{n,r}(\boldsymbol{\theta})\} = \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}\frac{\sin\left(\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|\right)}{\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|}\ \forall m,n, \tag{21}$$

  where A is the total surface of the walls and $\bar{\alpha}$ is the average absorption coefficient. If the reverberation time T_{60} is known, the average absorption coefficient can be approximated using Sabine's formula , i.e. $\bar{\alpha} = \frac{0.161V}{AT_{60}}$.

- A3 The direct and the reverberant components of the ATFs are uncorrelated (using spatial operator), i.e.,

$$\mathcal{E}_{\boldsymbol{\theta}}\{H_{m,d}(\boldsymbol{\theta})H^*_{n,r}(\boldsymbol{\theta})|d_m,d_n\} = 0,\ \forall m,n. \tag{22}$$

5.2. Spatially averaged output SNR of MWF

The objective of this subsection is to derive an analytical expression for the output SNR of the MWF using the statistical properties of the ATFs.

Using (18) and (19), the spatial characteristics ρ for each realization $\boldsymbol{\theta}$ is given by

$$\begin{aligned}\rho(\boldsymbol{\theta}) =& \mathbf{H}_d^H(\boldsymbol{\theta})\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\boldsymbol{\theta}) + \mathbf{H}_d^H(\boldsymbol{\theta})\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\boldsymbol{\theta}) \\ &+ \mathbf{H}_r^H(\boldsymbol{\theta})\mathbf{\Gamma}_v^{-1}\mathbf{H}_d(\boldsymbol{\theta}) + \mathbf{H}_r^H(\boldsymbol{\theta})\mathbf{\Gamma}_v^{-1}\mathbf{H}_r(\boldsymbol{\theta}).\end{aligned} \tag{23}$$

Without loss of generality, $\mathbf{H}_i(\boldsymbol{\theta})^H\mathbf{\Gamma}_v^{-1}\mathbf{H}_j(\boldsymbol{\theta})$ can be expressed as

$$\mathbf{H}_i^H(\boldsymbol{\theta})\mathbf{\Gamma}_v^{-1}\mathbf{H}_j(\boldsymbol{\theta}) = \sum_{m=1}^{M}\sum_{n=1}^{M}\check{\gamma}_{mn}H^*_{i,m}(\boldsymbol{\theta})H_{j,n}(\boldsymbol{\theta}), \tag{24}$$

where $\mathbf{H}_i(\boldsymbol{\theta})$ and $\mathbf{H}_j(\boldsymbol{\theta})$ can represent $\mathbf{H}_d(\boldsymbol{\theta})$ or $\mathbf{H}_r(\boldsymbol{\theta})$ and $\breve{\gamma}_{mn}$ represent the coefficients of the matrix $\boldsymbol{\Gamma}_v^{-1}$. Hence, $\rho(\boldsymbol{\theta})$ can be written as

$$\rho(\boldsymbol{\theta}) = \sum_{m=1}^{M}\sum_{n=1}^{M} \breve{\gamma}_{mn}\big(H^*_{d,m}(\boldsymbol{\theta})H_{d,n}(\boldsymbol{\theta}) + H^*_{d,m}(\boldsymbol{\theta})H_{r,n}(\boldsymbol{\theta}) + H^*_{r,m}(\boldsymbol{\theta})H_{d,n}(\boldsymbol{\theta}) + H^*_{r,m}(\boldsymbol{\theta})H_{r,n}(\boldsymbol{\theta})\big). \tag{25}$$

Using (22), the spatially averaged value of ρ given $\mathbf{d}$ (relative distance between source and microphones) is then equal to

$$\mathcal{E}_{\boldsymbol{\theta}}\{\rho(\boldsymbol{\theta})|\mathbf{d}\} = \sum_{m=1}^{M}\sum_{n=1}^{M} \breve{\gamma}_{mn}\big(\mathcal{E}_{\boldsymbol{\theta}}\{H^*_{d,m}(\boldsymbol{\theta})H_{d,n}(\boldsymbol{\theta})|\mathbf{d}\} + \mathcal{E}_{\boldsymbol{\theta}}\{H^*_{r,m}(\boldsymbol{\theta})H_{r,n}(\boldsymbol{\theta})|\mathbf{d}\}\big), \tag{26}$$

which, using (20) and (21), depends only on the relative distance between the source and the microphones and on the room properties ($A, \bar{\alpha}$), i.e.,

$$\boxed{\mathcal{E}_{\boldsymbol{\theta}}\{\rho(\boldsymbol{\theta})|\mathbf{d}\} = \sum_{m=1}^{M}\sum_{n=1}^{M} \breve{\gamma}_{mn}\Big(\frac{e^{j\frac{\omega}{c}(d_n-d_m)}}{(4\pi)^2 d_m d_n} + \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}\frac{\sin\left(\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|\right)}{\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|}\Big).} \tag{27}$$

For each realization of $\boldsymbol{\theta}$, (17) can be rewriten as

$$\mathrm{SNR_{out}}(\boldsymbol{\theta}) = \frac{\phi_s}{\phi_v}\rho(\boldsymbol{\theta}). \tag{28}$$

Hence, using (27), the spatially averaged output SNR, given the relative distance $\mathbf{d}$, is equal to

$$\boxed{\mathcal{E}_{\boldsymbol{\theta}}\{\mathrm{SNR_{out}}(\boldsymbol{\theta})|\mathbf{d}\} = \frac{\phi_s}{\phi_v}\sum_{m=1}^{M}\sum_{n=1}^{M} \breve{\gamma}_{mn}\Big(\frac{e^{j\frac{\omega}{c}(d_n-d_m)}}{(4\pi)^2 d_m d_n} + \frac{1-\bar{\alpha}}{\pi\bar{\alpha}A}\frac{\sin\left(\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|\right)}{\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|}\Big),} \tag{29}$$

which represents an analytical expression for the output SNR without having to measure or simulate the ATFs between source and microphones.

6. SIMULATION RESULTS

6.1. Experimental setup

In order to validate the theoretical results derived in the previous section, we consider the acoustical scenario depicted in Figure 2, which consists of $M = 6$ spatially distributed microphones in a room with dimensions 8 m×6 m×5 m and reverberation time $T_{60} = 0.25$ s (resulting in an average absorption coefficient $\bar{\alpha} = 0.65$). The circles in Figure 2 represent the microphone positions while the cross

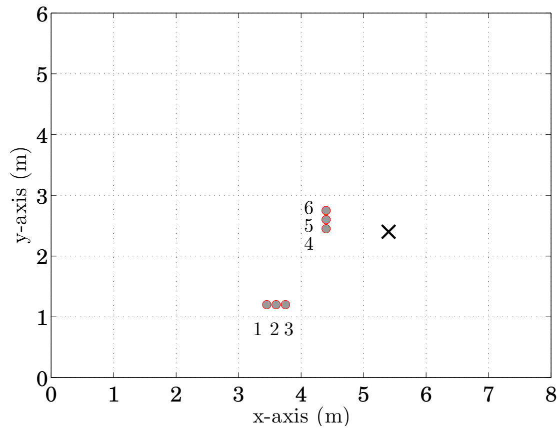

Fig. 2. Acoustical scenario with $M = 6$ spatially distributed microphones.

marker represents the position of the desired source such that $\mathbf{d} = [2.20\ \ 2.17\ \ 2.14\ \ 1.01\ \ 1.02\ \ 1.03]^T$. Two different subsets of microphones are considered. The first subset consists of the microphones # $1\cdots3$ and the second subset takes all microphones into account, i.e., microphones # $1\cdots6$. Different realizations of $\boldsymbol{\theta}$ given the constant relative distance $\mathbf{d}$ have been generated by rotating and translating the source-microphones configuration, and considering only the realizations of $\boldsymbol{\theta}$ that are located in the interior of the room and half a wavelength away from the walls. For each realization, impulse responses have been simulated using the image model [3], and the corresponding ATFs have been calculated. The length of the simulated impulse responses is 4096 samples and the sampling frequency $f_s = 16000$ Hz. Diffuse noise has been used and the noise coherence matrix was theoretically computed using

$$\gamma_{mn}(\omega) = \frac{\sin\left(\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|\right)}{\frac{\omega}{c}\|\mathbf{p}_m-\mathbf{p}_n\|}. \tag{30}$$

The average output SNR is numerically computed as

$$\overline{\mathrm{SNR_{out}}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\phi_s}{\phi_v}\rho(\tilde{\boldsymbol{\theta}}_i), \tag{31}$$

where N is the total number of realizations and $\tilde{\boldsymbol{\theta}}_i, i = 1\cdots N$ corresponds to a single realization of $\boldsymbol{\theta}$. Without loss of generality, the a priori input SNR is assumed to be frequency flat.

6.2. Results

For each subset of microphones, Figure 3 shows the average output SNR $\overline{\mathrm{SNR_{out}}}$ numerically computed using simulated ATFs by means of a Monte Carlo simulation with $N = 10000$ realizations of $\boldsymbol{\theta}$, together with the spatially averaged output SNR calculated analytically using (29). As can be seen, the spatially averaged output SNRs of the MWF computed using

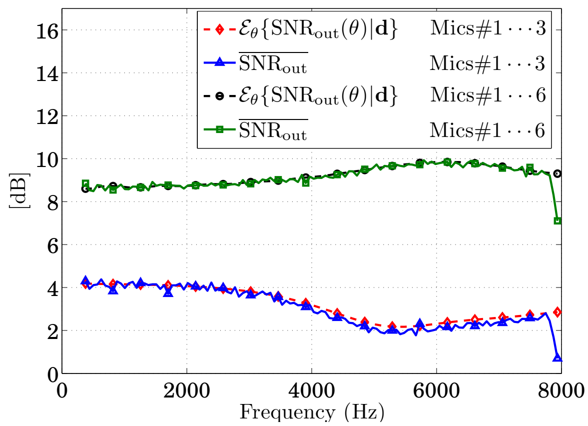

Fig. 3. Average output SNR calculated using a Monte Carlo simulation with 10000 realizations $\boldsymbol{\theta}$ and spatially averaged output SNR computed using statistical room acoustics.

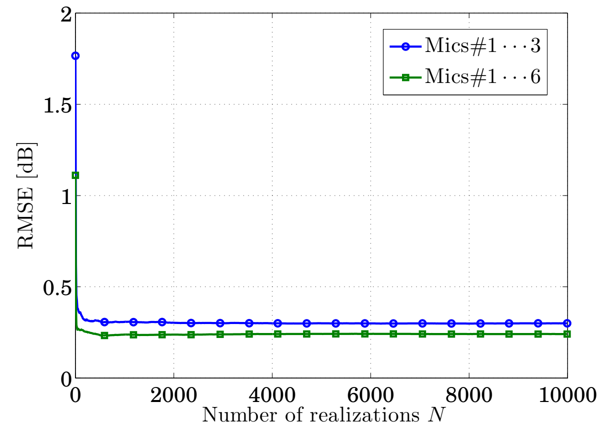

Fig. 4. Root mean square error between simulated and analytical results obtained using statistical room acoustics

statistical room acoustics are very close to the average output SNRs obtained using simulated ATFs. Therefore, if the source and microphone positions and the room characteristics (A,$\bar{\alpha}$) are known and if the noise coherence matrix can be estimated, the statistical properties of ATFs can be used to express the average output SNR of the MWF. Figure 4 depicts the root mean square error (RMSE) between the spatially averaged output SNR using SRA and the average output SNR computed using simulated ATFs as a function of the number of realizations N, i.e.,

$$\text{RMSE} = \sqrt{\sum_{\omega} \left| \mathcal{E}_{\boldsymbol{\theta}}\{\text{SNR}_{\text{out}}(\boldsymbol{\theta})|\mathbf{d}\} - \overline{\text{SNR}}_{\text{out}}(N) \right|^2}, \quad (32)$$

where $\mathcal{E}_{\boldsymbol{\theta}}\{\text{SNR}_{\text{out}}(\boldsymbol{\theta})|\mathbf{d}\}$ is computed using (29). As can be seen in Figure 4 when using the microphones #1 · · · 3, the RMSE is around 1.75dB for a single realization of $\boldsymbol{\theta}$, i.e. the spatially averaged output SNR given by (29) is not equal to the output SNR of the MWF for a single realization of $\boldsymbol{\theta}$. As expected, the larger the number of realizations, the smaller the RMSE between the analytical expression given by (29) and simulations using the image model. For a very large number of realizations, the RMSE converges to nearly zero. The fact that the RMSE does not converge exactly to zero, might be explained by imperfections of the image model. Although the spatially averaged output SNR does not express the output SNR of the MWF for a single realization $\boldsymbol{\theta}$, it can still be used to easily compare the performance of different subsets of microphones, given their relative source-microphones distances.

7. CONCLUSION

The theoretical performance of the MWF can be computed if the noise field and the ATFs between the desired source and the microphones are known. In this paper we have derived an analytical expression for the spatially averaged output SNR of the MWF using statistical room acoustics. This expression depends on the room dimensions, source-microphones distances and reverberation time. Simulation results have shown that the theoretical performance of the MWF computed using the statistical properties of ATFs is similar to the results obtained using simulated ATFs, providing an efficient way to compare the performance of different subsets of microphones.

8. REFERENCES

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

[2] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multi-channel Wiener filtering and Generalized Sidelobe Cancellation for multi-microphone noise reduction in hearing aid applications," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 4, pp. 487–503, Jul. 2005.

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

[4] M. R. Schroeder, "Frequency correlation functions of frequency responses in rooms," *Journal of the Acoustical Society of America*, vol. 34, no. 12, pp. 1819–1823, Dec. 1962.

[5] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment. Robustness results," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 3, pp. 311–319, Aug. 2002.

[6] M. Kompis, and N. Dillier, "Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. I. Prediction of the signal-to-noise-ratio improvement,"*Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1123–1133, Mar. 2001.